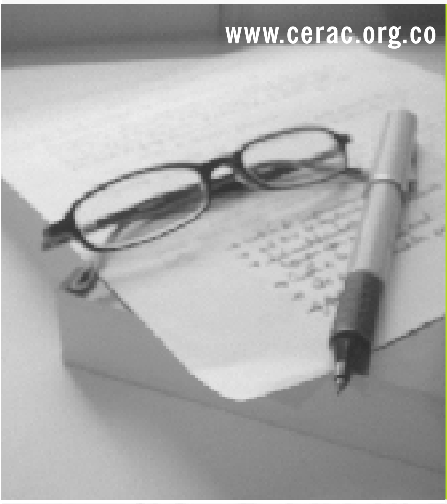
CERAC

Centro de Recursos para el Análisis de Conflictos



Documentos de CERAC

ISSN: 1909-1397

N° 7

Metodología para interpolar tamaños poblacionales

Mauricio Sadinle

Febrero, 2008

CERAC
Centro de Recursos para el Análisis de Conflictos

El Centro de Recursos para Análisis de Conflictos, CERAC, es un centro de investigación privado, especializado en el estudio de conflictos armados y violencia. Con sede en Bogotá, Colombia, CERAC fue creado en diciembre de 2004 por un grupo multidisciplinario de investigadores de diversas nacionalidades.

Entre los principales objetivos de CERAC se encuentra el estudio de la dinámica espacial y temporal de la violencia desde diferentes perspectivas metodológicas, haciendo énfasis en el impacto humano de los conflictos y el crimen.

La serie Documentos de CERAC busca presentar al público avances de investigación sobre análisis de conflicto desde una perspectiva académica.

Mauricio Sadinle está finalizando sus estudios de estadística en la Universidad Nacional de Colombia. También ha realizado estudios en economía en esta misma universidad. En la actualidad colabora en la construcción del paquete FactoClass de R que combina métodos factoriales y de clasificación. En CERAC trabaja en la estimación de tamaños poblacionales y en la caracterización de poblaciones.

The Conflict Analysis Resource Centre (CERAC) is a private research organization specialized in data-intensive studies of conflict and criminal violence. Based in Bogotá, Colombia, CERAC was established in December 2004 by an international and multidisciplinary group of researchers.

One of the main objectives of CERAC is the investigation of spatial and temporal dynamics of violence from a wide variety of methodological perspectives, emphasizing the human impact of conflict and crime.

The series Documentos de CERAC sought to disseminate preliminary research work on Conflict Analysis from a variety of academic methodological perspectives.

Mauricio Sadinle is finishing statistics at Universidad Nacional de Colombia. He has developed studies in economics at that same University. He contributes in the construction of the FactoClass R package, which combines factorial methods and classification. In CERAC he works in the estimation of size populations and in the characterization of populations.

ISSN: 1909-1397
CERAC - Centro de Recursos para el Análisis de Conflictos
Carrera 10 Nº 65-35 Of. 603/703
Teléfonos: +(571)3458572/3
Fax: +(571)2496117
www.cerac.org.co
Bogotá, Colombia
Cover Design: Adela Martínez Camacho

Metodología para interpolar tamaños poblacionales

Mauricio Sadinle*

Centro de Recursos para el Análisis de Conflictos - CERAC

Febrero de 2008

Resumen

Frecuentemente se tienen estimaciones del tamaño de una población para dos puntos del tiempo y se requieren estimaciones en puntos intermedios. En el presente documento se expone una solución sencilla trabajando con la tasa de variación media de la población en el periodo en consideración y se expone una propuesta para trabajar cuando hay segregación de poblaciones. Se presentan los resultados de varias aplicaciones para Bogotá y Colombia.

1. Desarrollo del método

En esta sección se sigue de cerca la presentación de Keyfitz & Caswell (2005) y de Preston, Heuveline & Guillot (2001). En demografía se llama *tasa* a la proporción entre las ocurrencias de un evento y la exposición a éste en un periodo de tiempo. Siguiendo esta idea, se tiene la siguiente definición general:

$$tasa[0,T] = \frac{\text{Número de ocurrencias entre 0 y T}}{\text{Suma de tiempos de exposición de los individuos entre 0 y T}}$$

Donde $[0,T]$ representa un periodo de tiempo de longitud T. Es claro que un individuo está expuesto a la ocurrencia del evento durante el tiempo que esté presente en la población dentro del periodo $[0,T]$, por lo que se toma como denominador la suma de tiempos de exposición de los individuos que estuvieron presentes en la población en algún momento entre 0 y T.

*Investigador. E-mail: mauricio.sadinle@cerac.org.co. Agradezco los comentarios de Jorge A. Restrepo y del equipo de trabajo del CERAC.

En este contexto los eventos de interés son los nacimientos, las muertes, las inmigraciones y las emigraciones, razones por las cuales puede variar el tamaño de una población. Para una variación del tiempo Δt se obtiene una variación en el tamaño de la población, $\Delta N(t) = N(t+\Delta t) - N(t)$, que refleja los cambios debidos a los cuatro eventos de interés anteriormente nombrados. Si se trata de calcular la tasa para el periodo comprendido entre un tiempo t y $t + \Delta t$, cuando Δt tiende a 0, es posible definir *la tasa de variación instantánea* como:

$$\begin{aligned} r(t) &= \lim_{\Delta t \to 0} \frac{N(t+\Delta t) - N(t)}{N(t)\Delta t} \\ &= \lim_{\Delta t \to 0} \frac{\Delta N(t)}{N(t)\Delta t} \\ &= \frac{1}{N(t)} \lim_{\Delta t \to 0} \frac{\Delta N(t)}{\Delta t} \\ &= \frac{1}{N(t)} \frac{dN(t)}{dt} \\ &= \frac{d \ln[N(t)]}{dt} \end{aligned}$$

Nótese que para la construcción de esta tasa se supone que todos los $N(t)$ individuos que están en la población en el tiempo t también están presentes hasta el tiempo $t + \Delta t$, lo cual es razonable teniendo en cuenta que Δt tiende a 0. Otro supuesto implícito en el desarrollo aquí presentado consiste en suponer la población como una función continua del tiempo.

Para encontrar una fórmula que permita estudiar el cambio poblacional entre dos puntos del tiempo 0 y T, se toma la integral de ambos lados de la ecuación anterior:

$$\begin{aligned} \int_0^T r(t)dt &= \int_0^T \frac{d \ln[N(t)]}{dt}\, dt \\ &= \ln N(t) \mid_0^T \\ &= \ln N(t) - \ln N(0) \\ &= \ln\left(\frac{N(t)}{N(0)}\right) \end{aligned}$$

Al tomar exponenciales a ambos lados se tiene:

$$e^{\int_0^T r(t)dt} = \frac{N(T)}{N(0)}$$

Por lo tanto:

$$N(T) = N(0)e^{\int_0^T r(t)dt} \tag{1}$$

De la definición del valor medio de una función para un intervalo, se tiene la *tasa de variación media de la población* en el periodo $[0, T]$ como:

$$\bar{r} = \frac{1}{T}\int_0^T r(t)dt$$

Es decir que:

$$\bar{r}T = \int_0^T r(t)dt$$

Al reemplazar en (1) se obtiene:

$$N(T) = N(0)e^{\bar{r}T} \tag{2}$$

Tomando logaritmo natural a ambos lados:

$$\begin{aligned} \ln N(T) &= \ln\left(N(0)e^{\bar{r}T}\right) \\ &= \ln N(0) + \bar{r}T \end{aligned}$$

De donde:

$$\begin{aligned} \bar{r} &= \frac{1}{T}\left(\ln N(T) - \ln N(0)\right) \qquad (3) \\ &= \frac{1}{T}\ln\left(\frac{N(T)}{N(0)}\right) \end{aligned}$$

Usualmente la formula obtenida en (2) se utiliza para hacer proyecciones poblacionales a periodos futuros, reemplazando T por un $t > T$. Lo anterior constituye una extrapolación del comportamiento promedio observado, pero también es de interés tener una estimación del tamaño en algún punto intermedio de $[0, T]$.

Para obtener una estimación para un tiempo t en $[0, T]$, se hace una interpolación utilizando los dos datos observados, y consiste en dividir este periodo en k subperiodos iguales, de tal manera que se pueda expresar t como $(T/k)i$, donde i es el indicador del subperiodo, que puede ser $0, 1, 2, ..., k$. Posteriormente se obtiene:

$$N(t) = N(0)e^{\bar{r}t} = N(0)e^{(\frac{\bar{r}T}{k})i} \tag{4}$$

2. Segregación poblacional

En ocasiones de la población de estudio se segrega una parte en algún momento del periodo en consideración y por lo tanto es necesario tener en cuenta la reducción del tamaño de la población original y la aparición de la nueva. En primera instancia se estudia el caso más sencillo de segregación y posteriormente se extiende la idea a situaciones más generales.

Sea A la población original y B la población segregada. Sea t' el momento en el que ocurre la segregación y $N_B(t)$ el tamaño de B. Nótese que $N_B(t)$ no está definido en $[0, t')$, pues antes de t', B estaba contenida en A. Usualmente no se conoce $N_A(t')$ ni $N_B(t')$, es decir que no se sabe cuál es el tamaño de la población que se segregó, pero se tiene $N_A(T)$ y $N_B(T)$, es decir, estimaciones de los tamaños de la población original y de la segregada en un tiempo posterior a la segregación.

De la fórmula (2) se obtiene:

$$N(0) = N(T)e^{-\bar{r}T}$$

Y por lo tanto, teniendo la tasa de variación promedio en un periodo $[0, T]$, es posible obtener dentro de éste estimaciones del tamaño poblacional con la información del final del periodo:

$$N(t) = N(T)e^{-\bar{r}(T-t)} \tag{5}$$

Que en términos de i, definido en la sección anterior, se expresa como:

$$N(t) = N(T)e^{\frac{\bar{r}T}{k}(i-k)} \tag{6}$$

En el manejo de la segregación, para obtener la tasa de variación promedio, se supone que toda la población crece homogéneamente, de tal manera que se puede calcular:

$$\bar{r} = \frac{1}{T}\ln\left(\frac{N_A(T) + N_B(T)}{N_A(0)}\right)$$

Y por lo tanto, se obtienen $N_A(t)$ y $N_B(t)$ de (5) y (6) para t en $[t', T]$:

$$\begin{aligned} N_A(t) &= N_A(T)e^{-\bar{r}(T-t)} = N_A(T)e^{\frac{\bar{r}T}{k}(i-k)} \\ N_B(t) &= N_B(T)e^{-\bar{r}(T-t)} = N_B(T)e^{\frac{\bar{r}T}{k}(i-k)} \end{aligned} \tag{7}$$

Y para los tiempos anteriores a la segregación, es decir t en $[0, t')$, se emplea la fórmula (4) para estimar $N_A(t)$.

La anterior presentación se puede extender a diversas situaciones. Suponga que una población C de tamaño N_C se conforma en un instante t' a partir de una parte de una población A de tamaño N_A y de otra población B de tamaño N_B. Bajo el supuesto de que A y B crecen a la misma tasa y homogéneamente, es posible calcular la tasa de variación promedio como:

$$\bar{r} = \frac{1}{T} \ln \left(\frac{N_A(T) + N_B(T) + N_C(T)}{N_A(0) + N_B(0)} \right)$$

Para los tiempos posteriores a la segregación de C, es decir t en $[t', T]$, la estimación de los tamaños poblacionales $N_A(t)$, $N_B(t)$ y $N_C(t)$ se puede realizar de manera análoga a (7), y para t en $[0, t')$, es decir tiempos anteriores a la segregación de C, $N_C(t)$ no está definido y se aplica la fórmula (4) para $N_A(t)$ y $N_B(t)$. En caso de que una nueva población se conforme a partir de partes de más de dos poblaciones, la idea anterior se extiende de manera natural.

Otra situación se presenta cuando se tienen, por ejemplo, tres segregaciones en el periodo $[0, T]$, es decir, de una población A de tamaño N_A se separa una población B de tamaño N_B en un instante t_B, una C de tamaño N_C en un instante t_C y una D de tamaño N_D en un instante t_D. Bajo el supuesto de que la población crece homogéneamente, se tiene:

$$\bar{r} = \frac{1}{T} \ln \left(\frac{N_A(T) + N_B(T) + N_C(T) + N_D(T)}{N_A(0)} \right)$$

Para t en el periodo $[t_D, T]$ la estimación de los tamaños poblacionales $N_A(t)$, $N_B(t)$, $N_C(t)$ y $N_D(t)$ se puede realizar de manera análoga a (7). Para el periodo $[t_C, t_D)$ la población D no estaba segregada y hacía parte de la población A, por lo que se tendría:

$$\begin{aligned}
N_A(t) &= [N_A(T) + N_D(T)]e^{-\bar{r}(T-t)} = [N_A(T) + N_D(T)]e^{\frac{\bar{r}T}{k}(i-k)} \\
N_B(t) &= N_B(T)e^{-\bar{r}(T-t)} = N_B(T)e^{\frac{\bar{r}T}{k}(i-k)} \\
N_C(t) &= N_C(T)e^{-\bar{r}(T-t)} = N_C(T)e^{\frac{\bar{r}T}{k}(i-k)}
\end{aligned}$$

Para el periodo $[t_B, t_C)$ la población C no estaba segregada y hacía parte de la población A, por lo que se tendría:

$$\begin{aligned}
N_A(t) &= [N_A(T) + N_D(T) + N_C(T)]e^{-\bar{r}(T-t)} = [N_A(T) + N_D(T) + N_C(T)]e^{\frac{\bar{r}T}{k}(i-k)} \\
N_B(t) &= N_B(T)e^{-\bar{r}(T-t)} = N_B(T)e^{\frac{\bar{r}T}{k}(i-k)}
\end{aligned}$$

Y para el periodo $[0, t_B)$, no se había producido ninguna segregación, por lo tanto se estima el tamaño de la población A utilizando la fórmula (4). La idea anterior se puede extender en caso de que dos o más de tres poblaciones se separen de A, teniendo cuidado con los tiempos en los que se presentan las segregaciones.

3. Aplicación

Ejemplo 1. Una aplicación del anterior método se presenta al emplear los datos de los censos colombianos de 1985 y del 2005 con el fin obtener estimaciones semestrales para este periodo. En este caso se tienen $T = 20$ años y por lo tanto se divide este periodo en $k = 20 * 2 = 40$, que sería el número de semestres. Aplicando la fórmula (3) se obtiene la tasa de variación media para este periodo:

$$\bar{r} = \frac{1}{20} \ln \left(\frac{42.888.592}{30.794.425} \right) = 0,0166$$

Y haciendo variar $i = 1, 2, ..., 40$ en (4) se obtienen las estimaciones deseadas. Las cifras se presentan redondeadas al entero más cercano:

Semestre	I			II		
Año	i	t	$N(t)$	i	t	$N(t)$
1985	0	0	30.794.425	1	0,5	31.050.517
1986	2	1	31.308.740	3	1,5	31.569.109
1987	4	2	31.831.644	5	2,5	32.096.362
1988	6	3	32.363.282	7	3,5	32.632.421
1989	8	4	32.903.799	9	4,5	33.177.433
1990	10	5	33.453.343	11	5,5	33.731.547
1991	12	6	34.012.065	13	6,5	34.294.916
1992	14	7	34.580.120	15	7,5	34.867.695
1993	16	8	35.157.661	17	8,5	35.450.039
1994	18	9	35.744.848	19	9,5	36.042.110
1995	20	10	36.341.843	21	10,5	36.644.069
1996	22	11	36.948.808	23	11,5	37.256.081
1997	24	12	37.565.910	25	12,5	37.878.315
1998	26	13	38.193.319	27	13,5	38.510.942
1999	28	14	38.831.206	29	14,5	39.154.134
2000	30	15	39.479.748	31	15,5	39.808.069
2001	32	16	40.139.120	33	16,5	40.472.925
2002	34	17	40.809.506	35	17,5	41.148.886
2003	36	18	41.491.088	37	18,5	41.836.136
2004	38	19	42.184.053	39	19,5	42.534.864
2005	40	20	42.888.592			

Ejemplo 2. Otro problema frecuente se presenta cuando se tienen proyecciones oficiales de la población por años y se quiere obtener estimaciones para periodos de tiempo más cortos. Por ejemplo, en la página de internet del Departamento Administrativo Nacional de Estadística de Colombia - DANE se encuentran las estimaciones

de población de Bogotá para los años 2006 y 2007 al 30 de junio en cada caso, que son 6.944.398 y 7.050.133 respectivamente. Es de interés obtener estimaciones del total poblacional de Bogotá para los trimestres intermedios, es decir 2006-III, 2006-IV y 2007-I. Tomando como tiempo 0 el 30 de junio de 2006 y dado que se está midiendo el tiempo en años, lo más razonable es tomar $T = 1$ el 30 de junio de 2007. En un año se encuentran 4 trimestres y por lo tanto se toma $k = 4$. Aplicando la fórmula (3) se obtiene $\bar{r} = \frac{1}{1} \ln \left(\frac{7.050.133}{6.944.398} \right) = 0{,}015$ y por lo tanto se tienen los siguientes resultados aplicando la fórmula (4):

Trimestre	i	t	$N(t)$
2006 - II	0	0	6.944.398
2006 - III	1	1/4	$N(0)e^{\frac{1}{4}\bar{r}} = 6.970.682$
2006 - IV	2	2/4	$N(0)e^{\frac{2}{4}\bar{r}} = 6.997.066$
2007 - I	3	3/4	$N(0)e^{\frac{3}{4}\bar{r}} = 7.023.549$
2007 - II	4	1	$N(0)e^{\frac{4}{4}\bar{r}} = 7.050.133$

Es importante recalcar que también es posible extrapolar el comportamiento promedio en este periodo utilizando la fórmula (4), tomando $t > T$ si se van a hacer proyecciones ó $t < 0$ si se van a estimar tamaños poblacionales para tiempos anteriores al primer dato con que se cuenta.

Ejemplo 3. La mayor parte de los municipios nuevos que aparecieron en Colombia entre 1993 y 2005 se segregaron de un solo municipio y por lo tanto en estos casos se aplica el método de estimación expuesto en la primera parte de la sección 2. Ejemplo de esto lo constituye la segregación del municipio de Coveñas del municipio de Santiago de Tolú, departamento de Sucre, en febrero de 2002. En este mismo departamento se presenta el segundo escenario de segregación con el municipio de El Roble, que se conforma definitivamente en marzo del 2000 con partes de los municipios de San Benito Abad, Sincé y Corozal. El tercer escenario de segregación presentado tiene lugar en el Chocó, con su capital Quibdó, de la cual se segregaron entre 1993 y 2005 tres municipios: Atrato, Medio Atrato y Rio Quito.

Las ideas presentadas en todo este documento se aplican a estimaciones mensuales para los municipios de Colombia en el periodo 1993-2005. Respecto de la segregación, en algunos casos fue necesario emplear combinaciones de las estrategias presentadas debido a que involucraban varios municipios iniciales y varias segregaciones. Las diferentes formas de interpolar empleadas se pueden seguir en las fórmulas de la hoja de cálculo que contiene las estimaciones. La información está disponible en:

http://www.cerac.org.co/interpoblacion/municipiosCol9305.xls

Bibliografía

APOSTOL, T. M. (**1988**). "Calculus", Vol. 1, Segunda edición, *Reverté*.

DANE (**2007**). "Ajuste municipal de la población 2005 y estimaciones por área 2006 y 2007"; `http://www.dane.gov.co/files/investigaciones/poblacion/proyepobla06_20`, (consultado el 22 de febrero de 2008).

DANE (**2008**). "Proyecciones de Población 2006 - 2007. Colombia"; `http://www.dane.gov.co/index.php?option=com_content&task=category&sectionid=23&id=477&Itemid=982`, (consultado el 22 de febrero de 2008).

DANE (**2008**). "Series de Población 1985 - 2020. Colombia"; `http://www.dane.gov.co/index.php?option=com_content&task=category&sectionid=16&id=496&Itemid=996`, (consultado el 22 de febrero de 2008).

KEYFITZ, N. & **CASWELL**, H. (**2005**). "Applied Mathematical Demography", Third edition, New York, *Springer*.

PRESTON, S. H., **HEUVELINE**, P. & **GUILLOT**, M. (**2001**). "Demography, Measuring and Modeling Population Processes"; *Blackwell Publishing*.