

## VALIDACIÓN DE LOS INDICADORES SINTOMÁTICOS EN LAS ESTIMACIONES POBLACIONALES: CENSO DEL 2000

*Edwin Chaves Esquivel*

Escuela de Matemática, Facultad de Ciencias Exactas y Naturales, Universidad Nacional, Heredia 86-3000, Costa Rica y Escuela de Estadística, Universidad de Costa Rica. Corel: echa@una.ac.cr

### ABSTRACT

Since the need of having good population estimates to lower administrative areas in the post census periods, it has been proposed the possibility of using auxiliary variables as simple alternative to produce indirect estimates. These variables have been called "Symptomatic indicators" of population change. This article researches about them using information got from the last population and housing census of 2000. It is analyzed the contingency of the produced estimates to cantonal level according to the census value. On the other hand it is discussed the coherence of these results with the ones got in a previous study that was done to census of 1984. The results showed that the combination of the indicators: electoral census, births, primary school registration, can bring right estimates even though there is a strong international immigration. There were used four methods that showed to be a good alternative to get these population estimates.

**Key words:** Small-Area Population estimates, Symptomatic indicators, Censal Ratio Methods, Difference of Rates, Ratio-Correlation Method, Correlation Rate.

### RESUMEN

Para contar con buenas estimaciones de población para áreas administrativas menores en los períodos poscensales, se planteó la posibilidad de utilizar variables auxiliares como una alternativa simple para producir estimaciones indirectas. A estas variables se les ha denominado "indicadores

sintomáticos" del cambio poblacional. La eficacia de estos indicadores se evaluó basándose en la información del censo de población y vivienda del año 2000. Se analiza la consistencia de las estimaciones producidas por cantones con respecto al valor censal. Además, se discute la coherencia de estos resultados con los del censo de 1984. Los resultados mostraron que la combinación de los indicadores: padrón electoral, nacimientos y matrícula escolar, pueden generar estimaciones acertadas aun en momentos de fuerte inmigración internacional.

### INTRODUCCIÓN

Las proyecciones y estimaciones de población son una herramienta fundamental para la planificación y programación de actividades en diferentes campos tales como: salud, educación, servicios públicos y privados, mercadeo, entre otros. Sin embargo, no siempre se cuenta con información poblacional de calidad cuando se requiere. Para los años censales se tienen estimaciones de mucha precisión; pero durante el período poscensal resulta sumamente difícil encontrar cifras sobre población que sean de calidad.

Debido a la problemática que conllevan esas estimaciones, fundamentalmente en áreas administrativas menores, desde hace décadas diversos investigadores han propuesto utilizar los indicadores sintomáticos, como una alternativa razonable al problema (Chaves 1996). Estos indicadores consisten en variables auxiliares que se encuentran fuertemente correlacionadas con el crecimiento poblacional de una localidad. De este modo, ellas pueden ser utilizadas para medir los cambios que se

producen en la población de una localidad. Las variables sintomáticas corresponden a información recolectada regularmente por empresas e instituciones públicas o privadas, para propósitos de orden administrativo; por ejemplo, la matrícula escolar, las declaraciones de impuestos, el número de abonados de los servicios básicos, el número de electores inscritos en el padrón electoral, entre otros (Serow y Rives 1995, Simpson *et al.* 1996).

La intención de recolectar estos datos no consiste en estimar poblaciones, por lo que generalmente se presentan varios problemas con su utilización.

Para que una fuente de información pueda considerarse como indicador sintomático de calidad debe satisfacer según Feeney *et al.* (1995) con: a) Sensibilidad: el indicador debe presentar una alta correlación con el tamaño poblacional. b) Disponibilidad: los datos deben estar actualizados periódicamente, mínimo año a año, para cada área menor. c) Confiabilidad: se debe entender la naturaleza de los datos que se van a utilizar. Es necesario tener claro ¿dónde y cómo fueron obtenidos? d) Estabilidad en definiciones de conceptos y en métodos de recolección: la definición de los conceptos y las técnicas de recolección no pueden variar significativamente durante los períodos de interés y e) Facilidad de obtención: las fuentes de información deben poder ser accesadas fácilmente y en forma oportuna. No deben tener limitaciones de confidencialidad.

Para el año 1996 se llevó a cabo un estudio para determinar los principales indicadores sintomáticos en Costa Rica, y precisar la calidad de las estimaciones que producían (Chaves 1996, 1997). El estudio consideró ocho fuentes de información: padrón electoral, nacimientos, defunciones, permisos de construcción de viviendas, número de asegurados de la CCSS, asistencia a centros de salud, abonados del servicio eléctrico y la matrícula en I y II ciclos de la enseñanza general básica y, además, se consideraron cuatro técnicas estadístico-matemáticas para trabajar con los indicadores sintomáticos, en la generación de estimaciones de población. Estos métodos se denominan: "razón censal", "diferencia de tasas", "razón correlación" y "tasa correlación" (Chaves 1997).

Según Chaves (1997), el padrón electoral, los nacimientos y la matrícula escolar, son buenos indicadores de los cambios en los tamaños de

población en las áreas administrativas de Costa Rica (provincias, cantones y distritos); mientras que las otras fuentes incumplieron al menos uno de los requisitos básicos informados por Feeney *et al.* (1995).

En este trabajo se pretende valorar la eficacia de estas técnicas como una alternativa para estimaciones confiables durante los períodos pos-censales o para evaluar otras fuentes de información.

## METODOLOGÍA

La evaluación de los procedimientos se basó en las estimaciones de población cantonal para el 30 de junio del año 2000 y se compararon con los datos del censo de población de dicho año; sin embargo, estas estimaciones se realizaron en un marco de inestabilidad migratoria, donde este componente demográfico ha sido determinante en los cambios poblacionales sufridos en muchos de los cantones del país en los últimos 10 años.

Los métodos utilizados en este estudio fueron:

### Método de Razón Censal

Se consideraron las tasas de ocurrencia de un determinado indicador sintomático. Supone que las áreas locales presentan una razón de cambio proporcional a la del área mayor, en el período comprendido entre el último censo y el momento de la estimación (Voss *et al.* 1995).

El procedimiento requiere determinar la tasa de ocurrencia de cada variable sintomática para las distintas áreas menores durante el último período censal, por lo que la población basada en la variable sintomática  $S$  para un cantón  $u$ ,  $t$  años después del último censo, se estima según la ecuación

$$r_o(u) = \frac{S_o(u)}{P_o(u)} \quad (1)$$

donde  $r_o(u)$  es la tasa de ocurrencia del indicador sintomático  $S$  para  $u$  en el año correspondiente al último censo.  $S_o(u)$  es la información de la variable sintomática  $S$  observada para  $u$  en el año correspondiente al último censo y  $P_o(u)$  es la población del cantón  $u$  medida en el último censo.

Posteriormente, se busca aproximar la tasa del indicador sintomático al momento  $t$ , por medio

de un parámetro  $\phi$ , el cual representa la razón de cambio en el indicador sintomático entre el año censal y el momento  $t$ , tal que:

$$r_t(u) = \phi r_o(u) \quad (2)$$

Para estimar el valor de  $\phi$  es necesario tener información del área mayor que contenga a  $u$ , para lo cual se pueden obtener estimaciones independientes de la población en el momento  $t$ . Esta área se denota con  $M$  y su población en el momento  $t$  por  $P_t(M)$ , donde la estimación de la población de  $M$  en  $t$  es  $\hat{P}_t(M)$ . De este modo, se calculan las tasas de ocurrencia de  $S$  en el área  $M$ , en el momento censal y en el momento  $t$ , utilizando la fórmula (1). Suponiendo que la razón de cambio en el indicador sintomático para el área mayor  $M$  puede aproximar efectivamente al valor de  $\phi$ , se tiene:

$$\hat{\phi} = \frac{\hat{r}_t(M)}{r_o(M)} \quad (3)$$

Entonces, la estimación de  $r_t(u)$  es:

$$\hat{r}_t(u) = \hat{\phi} r_o(u) = \frac{\hat{r}_t(M)}{r_o(M)} r_o(u) \quad (4)$$

Por lo tanto, la estimación poblacional para cada cantón  $u$ , en el momento  $t$ , viene dada por:

$$\hat{P}_t(u) = \frac{S_t(u)}{\hat{r}_t(u)} \quad (5)$$

Usualmente se obtiene más de una estimación de  $P_t(u)$  utilizando para ello diferentes indicadores sintomáticos. La estimación final es, generalmente, el promedio simple de las estimaciones individuales, aunque se han propuesto métodos para ponderarlas (Ericksen 1973, 1974).

La principal ventaja de este método es la simplicidad de aplicación. Presenta la desventaja que la dirección de las series sintomáticas podría estar afectada por factores distintos a los que justifican el cambio en la población. Esto significa que la razón para un cantón puede cambiar en una proporción diferente y realmente podría, incluso, moverse en dirección contraria a la razón del área mayor; de este modo, se incumpliría el supuesto básico del modelo.

## Método de Diferencia de Tasas

Este procedimiento es una variación del método de Razón Censal. La principal diferencia entre los dos métodos radica en la forma de estimar la tasa de ocurrencia del indicador sintomático para cada cantón. El método supone que el crecimiento de esta tasa, para el cantón  $u$ , es igual al crecimiento experimentado en ella para el área mayor  $M$ .

Considerando nuevamente el cálculo de estimaciones por medio del indicador sintomático  $S$ , la aproximación para la tasa de ocurrencia de este indicador en el momento  $t$ , para el cantón  $u$ , viene dada por la fórmula:

$$\hat{r}_t(u) = r_o(u) + [\hat{r}_t(M) - r_o(M)] \quad (6)$$

donde sus componentes se definen tal y como se especificó anteriormente.

## Método de Razón Correlación

En su forma básica, la aplicación de este método consiste en distribuir la población de un área mayor en sus cantones constituyentes. Esta técnica ha demostrado ser una de las más exactas en la estimación de población poscensal (Namboodiri 1972, Mandell y Taiman 1982, Feeney *et al.* 1995).

El supuesto básico del método supone una relación lineal entre las razones de cambio de las proporciones poblacionales y las razones de cambio de las proporciones de los indicadores, entre dos períodos. Considerando, además, que los cambios en el tiempo en esta relación afectan uniformemente todos los cantones, se obtiene un modelo más refinado que puede formularse como:

$$Y_u = a_0 + a_1 X_u \text{ donde}$$

$$Y_u = \frac{P_t(u)}{P_o(u)} \quad X_u = \frac{S_t(u)}{S_o(u)} \quad (7)$$

$$\frac{P_t(M)}{P_o(M)} \quad \frac{S_t(M)}{S_o(M)}$$

para todo cantón  $u$ . Los valores  $a_i$  ( $i = 0, 1$ ) corresponden a los coeficientes del modelo.

Por su parte,  $P$ ,  $S$ ,  $u$  y  $M$  están definidos tal como se indicó en los métodos anteriores y sus

subíndices 0 y 1 representan los momentos del penúltimo y último censos, respectivamente.

Esta ecuación indica que la razón de cambio en el tiempo de la proporción poblacional está sistemáticamente relacionada con la correspondiente razón de cambio en las proporciones de los indicadores. Se pueden introducir al modelo otros indicadores sintomáticos de manera que:

$$Y_u = a_0 + \sum_{j=1}^m a_j X_{u,j}$$

$$\text{donde: } X_{u,j} = \frac{\frac{S_{j,t}(u)}{S_{j,t}(M)}}{\frac{S_{j,o}(u)}{S_{j,o}(M)}} \quad (8)$$

para cada cantón  $u$  y cada indicador sintomático  $j$ .

Una vez que el modelo de regresión ha sido estimado resulta muy simple obtener las estimaciones de población para un período poscensal.

### Método de Tasa Correlación

Se le critica a la técnica Razón Correlación una inconsistencia en su lógica interna, ya que los coeficientes del modelo se obtienen tomando como referencia el tiempo entre los censos y, sin embargo, las estimaciones se efectúan en tiempos diferentes del período base, el cual es, por lo general, menor al tiempo entre los censos. Se ha propuesto una transformación al método de Razón Correlación, llamada "Tasa Correlación", que se basa en una aproximación exponencial de cambio, que se realiza tomando el logaritmo natural de cada razón y dividiéndolo entre el tiempo transcurrido entre los censos utilizados como referencia (Swanson y Tedrow 1984). Formalmente, el modelo se puede expresar por:

$$Y_u^* = a_0 + \sum_{j=1}^m a_j X_{u,j}^* \quad (9)$$

$$\text{donde: } Y_u^* = \frac{\ln(Y_u)}{k} \quad \text{y} \quad X_{u,j}^* = \frac{\ln(X_{u,j})}{k}$$

$Y_u$  y  $X_{u,j}$  están definidas tal como se indicó para la técnica de Razón Correlación y  $k$  es el intervalo de

tiempo transcurrido entre los censos, puede afectar la estabilidad natural de los coeficientes y las estimaciones producidas (Mandell y Taiman 1982, O'Hare 1980, Swanson 1980).

### Obtención de las estimaciones

De acuerdo con la definición de los métodos presentada anteriormente, para poder generar para el año 2000 estimaciones poblacionales por cantones, se requieren de los valores censales de los dos censos anteriores (1973 y 1984), además, de la información sintomática para esos años. La información de los censos y de los nacimientos fue obtenida de la página web "ccp.ucr.ac.cr" del Centro Centroamericano de Población de la Universidad de Costa Rica. Mientras que la información del padrón electoral y la matrícula escolar fue obtenida directamente del Tribunal Supremo de Elecciones y del Ministerio de Educación Pública.

Con esta información, mediante el uso de un software estadístico se procede a generar varias estimaciones de población para el año 2000. Estas estimaciones se comparan en contraposición con los valores censales y se determinan los errores producidos.

### Valoración de las estimaciones y análisis de los errores

Tal y como se efectuó en el estudio previo, para poder evaluar la bondad de las estimaciones producidas por los diferentes métodos, se determina el error porcentual ( $EP$ ) o porcentaje de desviación respecto al valor censal. Esta medida viene dada por la fórmula:

$$EP = \frac{(Pobl. estimada - Pobl. censal)}{Pobl. censal} \quad (10)$$

Un error porcentual positivo indica que la estimación sobreestimó la población censal, mientras que un valor negativo indica una subestimación.

Puesto que las estimaciones son prorrateadas para ajustar la población total del país, el error porcentual promedio toma valores cercanos a cero. Por tal razón, para analizar el comportamiento global de las estimaciones, se requiere calcular el valor absoluto del error porcentual ( $AEP$ ), determinado por la fórmula:

$$AEP = \left| \frac{\text{Pobl. estimada} - \text{Pobl. censal}}{\text{Pobl. censal}} \right| \cdot 100 \quad (11)$$

De este modo, el promedio del *AEP* fue una buena medida para evaluar la calidad de un conjunto de estimaciones, y un buen parámetro de comparación de los resultados de dos o más grupos de ellas.

Una vez analizados los errores producidos en las estimaciones se procede a compararlos con los obtenidos para 1984 en el estudio anterior. De este modo, se puede establecer la consistencia de los métodos y de los indicadores en los dos períodos considerados. Asimismo, es posible, determinar las características de las áreas que provocan los mayores errores.

## RESULTADOS

Dado que en el presente análisis se pretende estimar el tamaño de la población por medio de técnicas indirectas, con el auxilio de fuentes auxiliares de información, es prudente, primeramente, determinar el grado de correlación existente entre el tamaño real de la población con estas fuentes para los 81 cantones del país. Sin embargo, para medir de una mejor manera el comportamiento de esta correlación en el tiempo, el Cuadro 1 muestra

los coeficientes de correlación entre las razones de cambio de las magnitudes de los indicadores sintomáticos y de la población censal, para los períodos de 1973-1984 y 1984-2000. Esto permite observar en qué medida se relacionan los cambios en las magnitudes de los indicadores con respecto a los cambios en la población.

Claramente se puede establecer que para el período 1984-2000, los coeficientes de correlación de los indicadores con respecto a la población aumentan en forma importante en relación con lo ocurrido en el período 1973-1984. Estos resultados pueden hacer suponer que las estimaciones hechas para el año 2000 pueden tener mayor precisión que las realizadas para el año 1984.

Es importante señalar que debido a la inestabilidad que pueden presentar los nacimientos entre un año y otro, se procedió a trabajar con promedios trianuales, para estabilizar las cifras.

En primer lugar, se procede a emplear individualmente los indicadores sintomáticos para cada uno de los métodos de estimación. El Cuadro 2 presenta el análisis de los errores producidos en las estimaciones una vez comparadas en contraposición con la información censal, para los tres indicadores sintomáticos considerados. Además, se comparan con los resultados obtenidos en el análisis anterior.

**Cuadro 1**  
**Matrices de correlaciones entre las razones de cambio entre dos momentos censales, de tres indicadores sintomáticos y la población censal, períodos 1973-1984 y 1984-2000**

	Población	Padrón electoral	Nacimientos	Matrícula escolar
1973-1984*				
Población	1,00			
Pad. electoral	0,89	1,00		
Nacimientos	0,85	0,66	1,00	
Matr. escolar	0,79	0,81	0,48	1,00
1984-2000				
Población	1,00			
Pad. electoral	0,93	1,00		
Nacimientos	0,95	0,85	1,00	
Matr. escolar	0,93	0,83	0,90	1,00

\* Fuente: Chaves, Edwin (1997).

Cuadro 2

Comparación del valor absoluto en el error porcentual de las estimaciones poblacionales en los cantones por indicador sintomático, según el método utilizado, años 1984 y 2000

Método	1984*			2000		
	Pad. elec.	Nac.	Mat. esc.	Pad. elec.	Nac.	Mat. esc.
<b>Razón Censal</b>						
Promedio de AEP	5,0	8,8	11,3	6,4	6,7	7,8
Porcentaje de AEP inf. a 5%	61,8	42,6	30,9	45,7	49,4	37,0
Porcentaje de AEP sup. a 10%	13,2	33,8	51,5	13,6	24,7	30,9
<b>Diferencia de Tasas</b>						
Promedio de AEP	4,3	8,3	13,1	6,4	11,6	7,9
Porcentaje de AEP inf. a 5%	66,2	44,1	22,1	45,7	25,9	37,0
Porcentaje de AEP sup. a 10%	5,9	32,4	58,8	13,6	42,0	33,3
<b>Razón Correlación</b>						
Promedio de AEP	5,0	6,7	9,5	7,2	5,5	13,1
Porcentaje de AEP inf. a 5%	58,8	51,5	32,4	35,8	50,6	29,6
Porcentaje de AEP sup. a 10%	13,2	20,6	41,2	22,2	16,0	50,6
<b>Tasa Correlación</b>						
Promedio de AEP	4,9	6,4	9,6	6,9	5,6	7,4
Porcentaje de AEP inf. a 5%	55,9	51,5	36,8	38,3	51,9	40,7
Porcentaje de AEP sup. a 10%	11,8	23,5	42,6	21,0	18,5	29,6

\* Fuente: Chaves, Edwin (1997).

Los resultados señalan que el padrón electoral perdió consistencia para el año 2000 con respecto a lo ocurrido en 1984. Para las cuatro técnicas utilizadas, sus estimaciones aumentaron los errores en forma notoria. Esto puede tener explicación en que el padrón electoral es prácticamente insensible a la inmigración internacional, que como bien se ha dicho, ha tenido un importante aumento en los últimos 10 años. Sin embargo, la eficiencia de los otros dos indicadores ha aumentado entre estos dos años. Tanto el error promedio absoluto como el porcentaje de cantones con errores en las estimaciones mayores al 10% se han reducido para el año 2000, mientras que el porcentaje de estimaciones con errores menores del 5% ha aumentado ligeramente.

Con respecto a los métodos, la técnica llamada "tasa correlación" muestra cierta inestabilidad para el año 2000, específicamente al utilizar los nacimientos como indicador sintomático. Pareciera que para este indicador, el supuesto básico de esta técnica se incumple notoriamente en una buena parte de los cantones para el período 1984-2000. Al utilizar en forma individual los indicadores sintomáticos, las técnicas de regresión lineal ofrecen mejores resultados, aunque no son muy marcadas sus diferencias con respecto al método de "razón censal".

De acuerdo con los resultados anteriores, la utilización de más de un indicador a la vez puede influir en forma positiva en la reducción de los

errores en las mismas. Para los métodos de regresión, simplemente se incluyen en el modelo dos o más indicadores como variables predictoras. Mientras que para las técnicas de "razón censal" y "diferencia de tasas", se procede a promediar las estimaciones obtenidas con los indicadores individuales. El Cuadro 3 muestra las combinaciones de los indicadores que ofrecen la mayor precisión en las estimaciones, al utilizar dos o más indicadores.

El Cuadro 3 muestra las dos combinaciones de indicadores que proporcionaron los mejores resultados, aunque otras combinaciones tales como padrón electoral con matrícula escolar y nacimientos con matrícula escolar, ofrecieron resultados muy coherentes con los anteriores, pero con menor exactitud.

Existe una gran consistencia en los resultados para ambos años. Sin importar el método utilizado, las mejores estimaciones se obtienen cuando se combinan los tres indicadores sintomáticos. Este hecho se evidencia tanto para el año 1984 como

para el año 2000. Posiblemente, el efecto que la inmigración internacional produce a las estimaciones obtenidas por medio del padrón electoral se compensa con la intervención de los nacimientos y de la matrícula escolar, los cuales son sensibles a este tipo de migración. Sin embargo, la presencia del padrón electoral es fundamental, pues los nacimientos y la matrícula escolar por sí solos no producen la suficiente precisión.

En lo que respecta a las diferencias entre los métodos, la técnica denominada "diferencia de tasas" pareciera no tener la misma exactitud que las restantes, fundamentalmente en relación con los nacimientos, tal como se notó en el Cuadro 2. Los métodos de regresión lineal múltiple generan estimaciones, cuyos errores difícilmente superan el 10% y alrededor de un 70% de ellos son menores del 5%, estos resultados son muy coherentes para ambos años. El método de "razón censal" generó el menor error promedio y la menor desviación estándar para el año 2000; además, el 80% de los

**Cuadro 3**  
**Comparación del valor absoluto en el error porcentual y del error porcentual en las estimaciones poblacionales en los cantones por indicador sintomático, según el método utilizado, años 1984 y 2000**

Indicadores sintomáticos utilizados	Método							
	Razón Censal		Diferencia de Tasas		Razón Correlación		Tasa Correlación	
	1984	2000	1984	2000	1984	2000	1984	2000
<b>Padrón electoral y Nacimientos</b>								
Promedio de AEP	4,7	4,0	4,7	4,2	4,1	4,3	4,0	4,6
Desviación estándar de los EP	6,8	5,0	6,1	5,5	4,9	5,4	4,8	5,6
Valor máximo de AEP	21,6	11,7	20,7	20,0	12,1	13,8	12,5	14,7
Porcentaje de AEP inf. a 5%	66,2	69,1	64,7	59,3	72,1	63,0	72,1	63,0
Porcentaje de AEP sup. a 10%	10,3	3,7	10,3	2,5	2,9	6,2	2,9	7,4
<b>Padrón electoral, Nacimientos y Matrícula escolar</b>								
Promedio de AEP	4,8	3,4	5,4	4,5	3,8	3,7	3,8	3,7
Desviación estándar de los EP	5,9	4,4	6,6	6,1	4,7	4,6	4,6	4,6
Valor máximo de AEP	17,5	15,5	17,2	19,2	13,5	9,9	14,2	9,5
Porcentaje de AEP inf. a 5%	55,9	80,2	57,4	67,9	72,1	71,6	72,1	70,4
Porcentaje de AEP sup. a 10%	8,8	3,7	14,7	12,3	2,9	0,0	1,5	0,0

Fuente: Chaves, Edwin (1997).

cantones presentaron estimaciones con errores inferiores al 5%. Sin embargo, los resultados fueron muy diferentes con los ocurridos para el año 1984, donde la precisión no fue tan buena.

Dado que para este análisis se ha recurrido a métodos de regresión lineal múltiple, es necesario efectuar una reflexión con respecto al cumplimiento de los supuestos básicos de la técnica. Recordando en forma breve estos supuestos, la variable dependiente  $y_i$  debe estar normalmente distribuida con variancia constante  $\sigma^2$  y media  $\beta_0 + \beta_1 x_{i1} + \beta_2$

$x_{2i} + \dots + \beta_k x_{ki}$ , donde  $x_{ij}$  corresponde a las variables explicativas para  $i = 1, 2, \dots, n$  y  $j = 1, 2, \dots, k$ . Estos principios son equivalentes a indicar que los residuos en las predicciones del modelo se distribuyen normalmente con media cero y variancia constante  $\sigma^2$ . Además, los residuos no deben estar correlacionados con las variables predictoras. Por último, las variables explicativas  $x_{ij}$  no pueden estar correlacionadas entre sí (Johnson y Wichern 1988).

Para evaluar el cumplimiento de estos supuestos, en los modelos que proporcionaron las

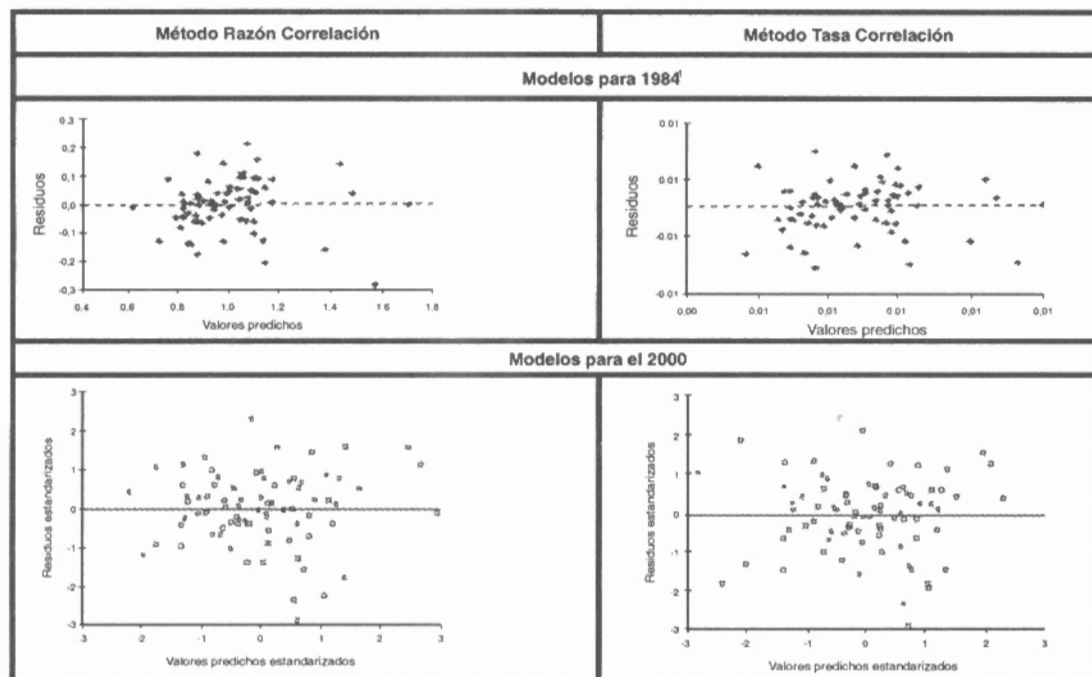
Cuadro 4

Resumen de los principales datos estadísticos para los modelos de regresión lineal múltiple, según el método aplicado con las variables padrón electoral, nacimientos y matrícula escolar

Modelo	R <sup>2</sup>	S <sup>2</sup>	F	Constante	Coefficientes b <sub>i</sub>	Significancia del b <sub>i</sub> (prueba t)
<b>Modelo utilizado en las estimaciones de 1984<sup>1</sup></b>						
Razón	0,82	0,008	98,4	0,15		
Correlación			(3,64) gl			
Pad. electoral					0,45	0,000
Nacimientos					0,23	0,040
Matr. escolar					0,16	0,070
Tasa	0,82	0,008	98,4	0,00		
Correlación			(3,64) gl			
Pad. electoral					0,45	0,000
Nacimientos					0,26	0,001
Matr. escolar					0,17	0,056
<b>Modelo utilizado en las estimaciones del 2000</b>						
Tasa	0,94	0,003	383,7	0,13		
Correlación			(3,77) gl			
Pad. electoral					0,31	0,000
Nacimientos					0,39	0,000
Matr. escolar					0,17	0,000
Razón	0,94	0,000	381,8	0,00		
Correlación			(3,77) gl			
Pad. electoral					0,34	0,000
Nacimientos					0,35	0,000
Matr. escolar					0,17	0,000

1. Fuente: Chaves, Edwin (1997).





**Figura 1. Figuras de residuos correspondientes a los modelos de regresión utilizados para estimar la población de los cantones de Costa Rica para los años 1984 y 2000.**

1. Fuente: Chaves, Edwin (1997).

mejores estimaciones, se muestra un análisis gráfico de residuos. Consiste en observar el comportamiento de los residuos en contraposición con los valores predichos, donde se espera obtener un comportamiento por completo aleatorio alrededor de la recta  $y = 0$  (Hernández y Bolaños 1993). Sin embargo, primeramente se muestra un análisis de los principales datos estadísticos de estos modelos. La información estadística se muestra en el Cuadro 4 y en la Figura 1 el análisis de los residuos.

La mayoría de resultados indican que los modelos utilizados para el año 2000 presentan una mayor consistencia, tienen menos variabilidad, mayor grado de explicación y más significancia estadística. En cuanto al análisis de residuos no se aprecian patrones importantes, especialmente en el modelo "tasa correlación". De acuerdo con este análisis y por la robustez de los modelos de regresión lineal, se puede deducir que es viable la utilización de estos modelos en las estimaciones de población por cantones en Costa Rica.

## DISCUSIÓN

Los resultados del estudio señalan que el empleo de los indicadores sintomáticos produce estimaciones de población muy consistentes en el tiempo. Existe una gran coherencia entre lo efectuado para el año 2000 con los resultados del estudio anterior para 1984. Inclusive, para el año 2000, a pesar que el tiempo transcurrido desde el último censo era aproximadamente 5 años mayor (11 años entre 1973 y 1984, 16 años entre 1984 y el 2000), las estimaciones fueron más precisas que las obtenidas antes. Esta situación se puede deber a la calidad de la información sintomática para este año, debido a que los procesos de recolección se han sistematizado y existe una mayor capacidad de almacenamiento de información.

Para el año 2000, el número de electores inscritos en el padrón electoral perdió exactitud en las estimaciones con respecto a 1984, lo cual es lógico, debido a su insensibilidad con respecto a la

inmigración internacional, la cual estuvo presente durante toda la década de los noventa y comienzos del siglo XXI. Sin embargo, las mejores estimaciones de población cantonal se producen cuando se complementa el uso del padrón electoral con los nacimientos y la matrícula escolar. Al combinar estos indicadores, se reducen los errores para todos los métodos, inclusive, se obtienen buenas estimaciones en los cantones con mayor inmigración internacional. Esta situación es digna de resaltar, pues en cantones de gran influencia inmigratoria como Sarapiquí, Limón, San José, Alajuela, o bien, cantones fronterizos como La Cruz, Upala, Los Chiles, Guatuso, Corredores, etc., se presentan errores relativamente bajos. Por el contrario, cantones como Tarrazú, Dota, Turrubares, San Isidro de Heredia, Aguirre, Montes de Oro y Talamanca, presentan los mayores errores. Esta circunstancia posiblemente se deba a problemas con la calidad de la información sintomática.

Debe resaltarse también que para el año 2000, las estimaciones se realizaron 16 años después del último censo de población, por lo que dadas las características de la metodología empleada, este hecho ejerce un efecto en su precisión. Sin embargo, prácticamente no se presentan estimaciones cuyos errores superen el 10% y en más del 70% de los cantones su población es estimada con un error menor del 5%. Esto sugiere que para períodos poscensales más cortos se pueden esperar estimaciones de mayor exactitud.

Las diferencias entre los cuatro métodos estudiados no son muy marcadas, no obstante, la técnica de "diferencia de tasas" muestra cierta inestabilidad principalmente con el uso de los nacimientos, pero ello repercute hacia las estimaciones hechas por combinación de indicadores. Por esta razón, se sugiere utilizar los otros tres métodos y comparar los resultados para analizar la consistencia de las predicciones efectuadas. En cuanto a los métodos que utilizan los modelos de regresión lineal, se pudo establecer que existe mucha consistencia desde el punto de vista estadístico en estos modelos, situación que se mantiene para los dos años analizados.

Finalmente, los resultados de este segundo estudio sobre el uso de los indicadores sintomáticos en las estimaciones de población para Costa Rica evidencian que estas técnicas pueden ponerse en práctica de una manera efectiva y simple. Su aplicación puede servir para generar estimaciones año a año, o bien, para comparar las proyecciones realizadas por otros métodos más sofisticados y que no se pueden estar replicando periódicamente. Sin embargo, debe recordarse que la utilización de las estimaciones no se puede realizar en forma indiscriminada sino que debe existir un proceso de reflexión y crítica de los datos, pues la exactitud de un método no es generalizada y por más robusta que parezca una técnica puede conducir a errores.

## REFERENCIAS

- Anónimo. 1988. Proyecciones Nacionales de Población 1975-2025. MIDEPLAN, CELADE y DGEC. San José, Costa Rica.
- Chaves, E. 1996. Indicadores sintomáticos en las estimaciones poblacionales a nivel cantonal. *Rev. Ciencias Económicas*. XVI (1):78-99.
- Chaves, E. 1997. Indicadores sintomáticos en las estimaciones poblacionales para áreas menores. Costa Rica. Tesis para optar al grado de Magister Scientiae, Universidad de Costa Rica. San José, Costa Rica.
- Erickson, E. P. 1973. A method for combining sample survey data and symptomatic indicators to obtain population estimates for local areas. *Demography* 10:137-160.
- Erickson, E. P. 1974. A regression method for estimating population changes of local areas. *J. Amer. Statistical Assoc.* 69:867-875.
- Feeney, D., J. Hibbs & T. Gillaspay. 1995. Ratio-Correlation Method. *In*: Rives, N., W. Serow y A. Lee. Basic methods for preparing small-area population estimate. Applied population laboratory, Department of Rural Sociology, College of Agricultural and Life Sciences. University of Wisconsin-Madison/Extension. Págs. 118-136.
- Hernández, O. & R. Bolaños (1993). SPSS/Pc + básico. EUCR. San José, Costa Rica.
- Johnson, R. & D. Wichern. 1988. Applied Multivariate Statistical Analysis. Segunda edición. Prentice Hall, New Jersey. University of Wisconsin. Págs. 273-339.
- Mandell, M. & J. Taiman. 1982. Measuring temporal stability in regression models of population estimation. *Demography* 19:135-146.
- Namboodiri, K. 1972. On the ratio-correlation and related method of subnational population estimation. *Demography* 9:443-453.
- O'Hare, W. 1980. A note on the use of regression methods in population estimates. *Demography* 17:341-343.
- Rives, N. W. 1982. Assessment of a survey approach, págs. 79-96. *In*: Lee, E. S. and H. F. Goldsmith (eds.). Population estimates: Methods for small area analysis. SAGE Publications. Beverly Hills, California.
- Serow, W. & N. Rives. 1995. Small area analysis: Assessing the state of the art. Presentado *In*: Rives, N., W. Serow y A. Lee. Basic methods for preparing small-area population estimate. Applied population laboratory, Department of Rural Sociology, College of Agricultural and Life Sciences. University of Wisconsin-Madison/Extension. Págs. 1-9.
- Simpson, S. *et al.* 1996. Updating Small Area Population Estimates in England and Wales. *J. Am. Statistical Assoc.* 159:235-247.
- Swanson, D. 1980. Improving accuracy in multiple regression estimates of population using principles from causal modelling. *Demography* 17:413-427.
- Swanson, D. & L. Tedrow. 1984. Improving the measurement of temporal change in regression models used for county population estimates. *Demography* 21:373-381.
- Voss, P., K. Palit & Krebs. 1995. Censal ratio methods. *In*: Rives, N., W. Serow y A. Lee. Basic methods for preparing small-area population estimate. Applied population laboratory, Department of Rural Sociology, College of Agricultural and Life Sciences. University of Wisconsin-Madison/Extension. Págs. 71-88.