

## **Validación de los indicadores sintomáticos en las estimaciones poblacionales: censo del 2000**

*Edwin A. Chaves Esquivel<sup>1</sup>*

### **RESUMEN**

Ante la necesidad de contar con buenas estimaciones de población para áreas administrativas menores en los períodos postcensales, se ha planteado 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. En este artículo se investiga su uso apoyándose en la información del último censo de población y vivienda del año 2000. Se analiza la consistencia de las estimaciones producidas a nivel cantonal con respecto al valor censal. Además se discute la coherencia de estos resultados con los obtenidos en un estudio anterior efectuado para el censo de 1984.

Los resultados mostraron que la combinación de los indicadores: padrón electoral, nacimientos y matrícula escolar, pueden generar acertadas estimaciones aun en momentos de fuerte inmigración internacional. Para ello se aplicaron cuatro métodos que mostraron ser una buena alternativa para obtener estas estimaciones de población.

### **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, etc. Sin embargo, no siempre se cuenta con información poblacional de calidad cuando se le requiere. Para los años censales se tienen estimaciones de mucha precisión; pero durante el período postcensal, resulta sumamente difícil encontrar cifras sobre población que sean de calidad. Debido a la problemática que conllevan estas estimaciones, fundamentalmente a nivel de áreas administrativas menores, desde hace muchas 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 al 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, etc. (Serow y Rives, 1995; Simpson y otros, 1996).

Es claro, que 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 (Feeney, Hibbs y Gillaspay, 1995):

---

<sup>1</sup> Máster en estadística con énfasis en población, profesor Universidad Nacional y Universidad de Costa Rica, investigador asociado del Centro Centroamericano de Población, echaves@ccp.ucr.ac.cr

- 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.
- 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 a nivel nacional, y precisar la calidad de las estimaciones que producían (Chaves 1996, Chaves 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. Además se consideraron cuatro técnicas estadístico-matemáticas para trabajar con los indicadores sintomáticos, en la generación de estimaciones de población. Estos métodos se denominan: Método de “razón, censal”, método de “diferencia de tasas”, método de “razón correlación y método de “tasa correlación” (Chaves 1997). Los resultados señalaron que solamente tres de las fuentes cumplían a cabalidad para ser utilizados como indicadores de los cambios en los tamaños de población a nivel de las áreas administrativas: provincias, cantones y distritos. En orden de importancia los indicadores seleccionados fueron: El padrón electoral, los nacimientos y la matrícula escolar. Las restantes fuentes incumplían al menos uno de los requisitos básicos (Chaves 1997).

Por las características de los métodos de estimación utilizados, se necesitó información sobre los censos de 1963 y 1973, así como los datos sintomáticos de esos años. Las estimaciones fueron validadas con respecto al censo de 1984. A pesar de utilizar información tan antigua, los resultados en las estimaciones fueron aceptables; pero quedó la incertidumbre sobre la validez de los indicadores y de los métodos utilizados para otros períodos de tiempo.

Por esta razón, el presente trabajo busca legitimar los resultados obtenidos en el estudio anterior. De este modo, se pretende ofrecer estas técnicas como una alternativa para contar con estimaciones aceptables durante los períodos postcensales o para evaluar otras fuentes de información. La evaluación de los procedimientos se realiza determinando estimaciones de población a nivel cantonal para el 30 de junio del año 2000 y contrastado esos resultados con los datos obtenidos en el censo de población de dicho año. Para estas estimaciones se cuenta con información sintomática más reciente y confiable que la utilizada en el estudio anterior, por lo que se podría suponer que las estimaciones serían de mayor calidad. Sin embargo; estas estimaciones se realizan 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 para los últimos 10 años.

## METODOLOGÍA

A continuación se procede a efectuar una descripción de las técnicas utilizadas para producir estimaciones mediante el uso de los indicadores sintomáticos. Los cuatro métodos empleados requieren del conocimiento sobre el tamaño de la población en el área mayor que las comprende o de una buena estimación de ésta (población nacional). Esto quiere decir que no producen estimaciones por si solos sino que realizan una distribución de la población total entre las distintas áreas menores que la conforman. Esto obliga a que las estimaciones obtenidas deban prorratearse es decir el conjunto de ellas debe reproducir la población total.

### Método de Razón Censal

Este método considera 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, Palit, Kale y Krebs; 1995).

En primer lugar, el procedimiento requiere determinar la tasa de ocurrencia de cada variable sintomática para las distintas áreas menores durante el último período censal. Supóngase que se desea estimar la población basada en la variable sintomática  $S$  para un cantón  $u$ ,  $t$  años después del último censo. Se calcula:

$$r_0(u) = \frac{S_0(u)}{P_0(u)} \quad (1)$$

$r_0(u)$  es la tasa de ocurrencia del indicador sintomático  $S$  para  $u$  en el año correspondiente al último censo.

$S_0(u)$  es la información de la variable sintomática  $S$  observado para  $u$  en el año correspondiente al último censo.

$P_0(u)$  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:

$$r_t(u) = \phi r_0(u) \quad (2)$$

Para estimar el valor de  $\phi$  es necesario tener información del área mayor que contenga a  $u$ , para la 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  $M$ , al momento censal y al 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:

$$\hat{\phi} = \frac{\hat{r}_t(M)}{r_0(M)} \quad (3)$$

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

$$\hat{r}_t(u) = \hat{\phi} r_0(u) = \frac{\hat{r}_t(M)}{\hat{r}_0(M)} r_0(u) \quad (4)$$

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

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

Usualmente se obtiene más de una estimación de  $\hat{P}_t(u)$  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; Ericksen, 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:

$$\hat{r}_t(u) = r_0(u) + [\hat{r}_t(M) - r_0(M)] \quad (6)$$

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

### Método de Razón Correlación

En su forma básica, la aplicación del método consiste en distribuir la población de un área mayor en sus cantones constituyentes. Esta técnica ha demostrado ser uno de los más exactos en la estimación de población postcensal (Namboodiri, 1972; Mandell y Tayman, 1982; Fenney, Hibbs y Guillaspy, 1995).

El supuesto básico del método asume 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:

$$Y_u = a_0 + a_1 X_u \quad (7)$$

$$\text{donde: } Y_u = \frac{\frac{P_1(u)}{P_1(M)}}{\frac{P_0(u)}{P_0(M)}} \quad \text{y} \quad X_u = \frac{\frac{S_1(u)}{S_1(M)}}{\frac{S_0(u)}{S_0(M)}} \quad \text{para todo cantón } u$$

$a_i$  ( $i = 0,1$ ) corresponde a los coeficientes del modelo

Los valores  $P$ ,  $S$ ,  $u$  y  $M$  están definidos tal como se indicó en los métodos anteriores y sus subíndices 0 y 1 representan los momentos censales del penúltimo e último censo 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:

$$Y_u = a_0 + \sum_{j=1}^m a_j X_{u,j} \quad \text{donde} \quad X_{u,j} = \frac{\frac{S_{j,1}(u)}{S_{j,1}(M)}}{\frac{S_{j,0}(u)}{S_{j,0}(M)}} \quad (8)$$

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 postcensal.

### El 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. Esto puede afectar la estabilidad natural de los coeficientes y las estimaciones producidas (Mandell y Tayman, 1982; O'Hare, 1980; Swanson, 1980).

Se ha propuesto una transformación al método de Razón Correlación, llamada "Tasa Correlación", que se basa en una aproximación exponencial de cambio, y que se realiza tomando el logaritmo natural de cada razón y dividiéndola 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 definidos tal como se indicó para la técnica de Razón Correlación y  $k$  es el intervalo de tiempo transcurrido entre los censos.

### Obtención de las estimaciones

De acuerdo con la definición de los métodos presentada anteriormente, para poder generar para el año 2000 estimaciones poblacionales a nivel cantonal, se requiere 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 “<http://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 contra 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:

$$EP = \frac{(Pobl. estimada - Pobl. censal)}{Pobl. censal} \cdot 100 \quad (10)$$

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:

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

De este modo, el promedio de los  $AEP$  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 No. 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 censos de 1973-1984 y 1984-2000. Esto permite observar en que medida se relacionan los cambios en las magnitudes de los indicadores con respecto a los cambios en la población.

**Cuadro No. 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**

	<b>Población</b>	<b>Padrón Electoral</b>	<b>Nacimientos</b>	<b>Matrícula escolar</b>
<b>1973-1984<sup>1</sup></b>				
<b>Población</b>	1,00			
<b>Pad. Electoral</b>	0,89	1,00		
<b>Nacimientos</b>	0,85	0,66	1,00	
<b>Matr. Escolar</b>	0,79	0,81	0,48	1,00
<b>1984-2000</b>				
<b>Población</b>	1,00			
<b>Pad. Electoral</b>	0,93	1,00		
<b>Nacimientos</b>	0,95	0,85	1,00	
<b>Matr. Escolar</b>	0,93	0,83	0,90	1,00

<sup>1</sup> **Fuente:** Chaves, Edwin (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

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 con respecto a 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 No. 2 presenta el análisis de los errores producidos en las estimaciones una vez comparadas contra la información censal, para los tres indicadores sintomáticos considerados. Además se comparan con los resultados obtenidos en el análisis anterior.

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 amentado 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% a aumentado ligeramente.

**Cuadro No. 2. Comparación del valor absoluto en el error porcentual de las estimaciones poblacionales a nivel cantonal por indicador sintomático según método utilizado, 1984 y 2000.**

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

<sup>1</sup> Fuente: Chaves, Edwin (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

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 individualmente 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 positivamente en la reducción de los errores en las mismas. Para los métodos de regresión,



simplemente se incluye 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 No. 3 muestra la combinación de los indicadores que ofrecen la mayor precisión en las estimaciones, al utilizar dos o más indicadores.

**Cuadro No. 3. Comparación del valor absoluto en el error porcentual y del error porcentual en las estimaciones poblacionales a nivel cantonal por indicador sintomático según método utilizado, años 1984 y 2000**

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

<sup>1</sup> Fuente: Chaves, Edwin (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

El cuadro 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 gran consistencia en los resultados obtenidos para ambos años. Sin importar el método, 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 si 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 con respecto a los nacimientos, tal como se notó en el Cuadro No. 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 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 brevemente estos supuestos, la variable dependiente  $y_i$  debe estar normalmente distribuida con variancia constante  $\sigma^2$  y media  $\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki}$ , donde  $x_{ij}$  corresponden 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. Finalmente las variables explicativas  $x_{ij}$  no pueden estar correlacionadas entre sí (Jonson y Wichern, 1988).

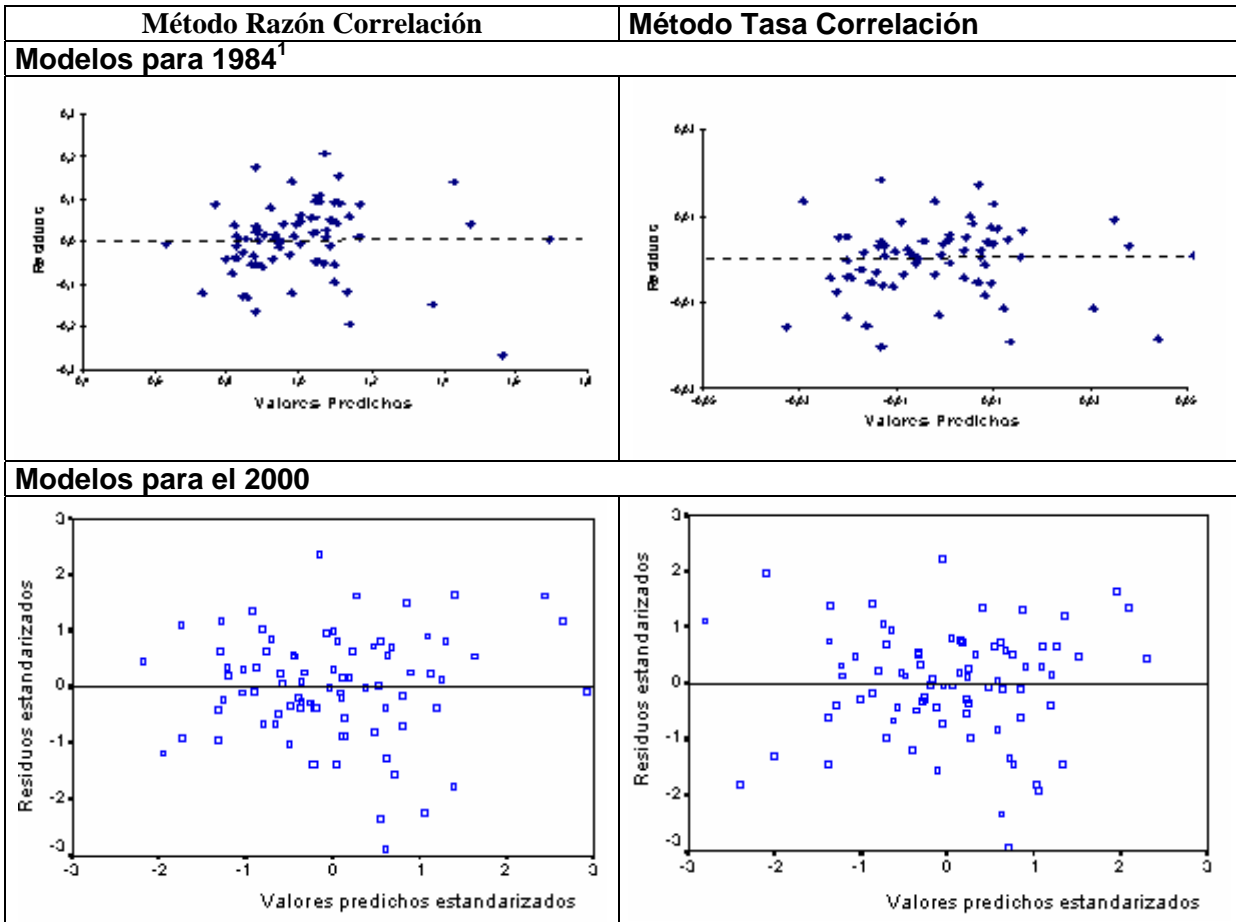
Para evaluar el cumplimiento de estos supuestos, en los modelos que proporcionaron las mejores estimaciones, se muestra un análisis gráfico de residuos. Consiste en observar el comportamiento de los residuos contra los valores predichos, donde se espera obtener un comportamiento completamente aleatorio alrededor de la recta  $y = 0$  (Hernández y Bolaños, 1993). Sin embargo; primeramente se presenta un análisis de los principales datos estadísticos de estos modelos. La información estadística se presenta en el Cuadro No. 4. y en el Gráfico No. 1 se presenta el análisis de los residuos

**Cuadro No. 4. Resumen de los principales estadísticos para los modelos de regresión lineal múltiple según el método aplicado con la variables padrón electoral, nacimientos y matrícula escolar.**

Modelo	$R^2$	$S^2$	$F$	Constante	Coefficiente $s$ $b_i$	Significancia del $b_i$ (prueba t)
<b>Modelo utilizado en las estimaciones de 1984<sup>1</sup></b>						
<b>Razón Correlación</b>	0,82	0,008	98,4 (3,64) gl	0,15		
Pad. Electoral					0,45	0,000
Nacimientos					0,23	0,040
Matr. Escolar					0,16	0,070
<b>Razón Correlación</b>	0,82	0,008	98,4 (3,64) gl	0,00		
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>						
Modelo	$R^2$	$S^2$	$F$	Constante	Coefficiente $s$ $b_i$	Significancia del $b_i$ (prueba t)
<b>Razón Correlación</b>	0,94	0,003	383,7 (3,77) gl	0,13		
Pad. Electoral					0,31	0,000
Nacimientos					0,39	0,000
Matr. Escolar					0,17	0,000
<b>Razón Correlación</b>	0,94	0,000	381,8 (3,77) gl	0,00		
Pad. Electoral					0,34	0,000
Nacimientos					0,35	0,000
Matr. Escolar					0,17	0,000

<sup>1</sup> Fuente: Chaves, Edwin (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

**Gráfico No 1. Gráfico 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, según método**



<sup>1</sup> Fuente: Chaves, Edwin (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

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 pueden deducir que es viable la utilización de estos modelos en las estimaciones de población a nivel cantonal en Costa Rica.

## DISCUSIÓN

Los resultados del estudio señalan que el empleo de los indicadores sintomáticos producen estimaciones de población muy consistentes en el tiempo. Existe 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 en 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. Esto 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 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 de siglo. Sin embargo; las mejores estimaciones de población a nivel cantonal, se producen cuando se complementa el uso del padrón electoral junto a los nacimientos y a 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. Esto es digno de resaltar pues en cantones de gran influencia inmigratoria como Sarapiquí, Limón, San José, Alajuela o cantones fronterizos como La Cruz, Upala, Los Chiles, Guatuso, Corredores, etc., se presentan errores relativamente bajos. Por el contrario, cantones como Tarrazú, Dota, Turubares, San Isidro de Heredia, Aguirre, Montes de Oro y Talamanca son los que presentan los mayores errores, posiblemente debido 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 postcensales 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 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 pueden realizar indiscriminadamente 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.

**BIBLIOGRAFÍA**

- CELADE, DGEC y MIDEPLAN 1988. Proyecciones Nacionales de Población 1975-2025 MIDEPLAN, CELADE y DGEC. San José, Costa Rica.
- Chaves, Edwin (1996). Indicadores Sintomáticos en las estimaciones poblacionales a nivel cantonal. Revista de Ciencias Económicas. Vol. XVI. No. 1, pp.78-99.
- Chaves, Edwin (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
- Ericksen, Eugene P (1973). A method for combining sample survey data and symptomatic indicators to obtain population estimates for local areas. Demography, 10, 137-160.
- Ericksen, Eugene P (1974). A regression method for estimating population changes of local areas. Journal of the American Statistical Association, 69, 867-875.
- Feeney, Donald; Hibbs, James y Gillaspay, Thomas (1995). Ratio-Correlation Method. Presentado en Rives, Norfleet y otros Basic Methods for Preparing Small-Area Population Estimate (pag. 118-136). Applied Population Laboratory, Department of Rural Sociology, College of Agricultural and Life Sciences. University of Wisconsin-Madison/Extension.
- Johnson, Richard y Wichern, Dean (1988) Applied Multivariate Statistical Analysis. Segunda Edición. Prentice Hall, New Jersey. University of Wisconsin.
- Mandell, Marylou y Tayman, Jeffrey (1982). Measuring Temporal Stability in Regression Models of Population Estimation. Demography, 19, 135-146.
- Namboodiri, Krishnan (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, Norfleet W (1982). Assessment of a Survey Approach. En Lee, Everetts S y Goldsmith, Harold F. Population Estimates: Methods for small area analysis (pag. 79-96). SAGE Publications. Beverly Hills, California.
- Serow, William y Rives, Norfleet (1995). Small Area Analysis: Assessing the State of the Art. Presentado en Rives, Norfleet y otros Basic Methods for Preparing Small-Area Population Estimates (pág. 1-9). Applied Population Laboratory, Department of Rural Sociology, College of Agricultural and Life Sciences. University of Wisconsin-Madison/Extension.
- Simpson, Stephen y otros (1996). Updating Small Area Population Estimates in England and Wales. Journal of the American Statistical Association, 159, 235-247
- Swanson, David (1980). Improving Accuracy in Multiple Regression Estimates of Population Using Principles From Causal Modelling. Demography, 17, 413-427.

Swanson, David y Tedrow, Lucky (1984). Improving the Measurement of Temporal Change in Regression Models Used for County Population Estimates. Demography, 21,373-381.

Voss, Paul R. y otros (1995). Censal Ratio Methods. En Rives, Norfleet y otros, Basic Methods for Preparing Small-Area Population Estimate (pag. 71-88). Applied Population Laboratory, Department of Rural Sociology, College Agricultural and Life Sciences. University of Wisconsin-Madison/Extension.

Wolter, Kirk y Causey, Beverley. Evaluation of Procedures For Improving Population Estimates for Samall Areas. Journal of the American Statistical Association, 86, 278-284

## ANEXO: EJEMPLOS DE APLICACIÓN DE LOS MÉTODOS

Para que el lector pueda tener una mejor comprensión en la aplicación de las técnicas de estimación poblacional, este anexo presenta un ejemplo de cálculo. Se efectuaron estimaciones de la población del cantón Central de San José, utilizando en cada caso la información del padrón electoral.

**Método de Razón Censal:** Para analizar un ejemplo de esta técnica, se debe contar con la tasa de electores para cada cantón y para Costa Rica en 1973. Además, un estimado de la tasa de electores para Costa Rica en 1984. De acuerdo con la información proporcionada por el Cuadro A 1.3 del Anexo 1, para el cantón Central de San José se tiene:

	Electores inscritos			Población censal		
	1973	1984	2000	1973	1984	2000
<b>San José</b>	120 647	179 237	209 260	215 441	241 464	309 672
<b>Costa Rica</b>	791 033	1 374 278	2 163 130	1 871 780	2 416 809	3 810 179

Entonces, si se representa con  $r$  la tasa de electores inscritos, se tiene que:

$$r(\text{San José})_{1984} = \frac{179\,237}{241\,464} = 0,74229$$

$$r(\text{Costa Rica})_{1984} = \frac{1\,374\,278}{2\,416\,809} = 0,56863$$

$$r(\text{Costa Rica})_{2000} = \frac{2\,163\,130}{3\,810\,179} = 0,56772$$

La técnica supone que la tasa de electores de San José para el 2000, viene dada por:

$$r(\text{San José})_{2000} = f r(\text{San José})_{1984}$$

donde el valor de  $f$ , puede ser estimado por:

$$\hat{f} = \frac{r(\text{Costa Rica})_{2000}}{r(\text{Costa Rica})_{1984}} = \frac{0,56772}{0,56863} = 0,99840$$

Con lo cual, la estimación de  $r(\text{San José})_{2000}$  es:

$$\hat{r}(\text{San José})_{2000} = 0,99840 \cdot 0,74229 = 0,74110$$

La estimación simple de la población para el cantón central de San José en el 2000 viene dada por:



$$\hat{P}(\text{San José})_{2000} = \frac{\text{Elec. Insc.}_{2000}}{\hat{r}(\text{San José})_{2000}} = \frac{209\,260}{0,74110} \approx 282\,364$$

Este proceso se efectuó para los 81 cantones, por lo que la suma de las estimaciones poblacionales de todos ellos debe coincidir con la población total del país (3 810 179). Sin embargo, la suma de las estimaciones simples para todos los cantones es 3 879 263. Por esta razón es necesario prorratear las estimaciones, de manera que la suma de ellas coincida con la población censal de Costa Rica para 2000. Para este fin, la estimación de cada cantón se multiplica por el cociente:

$$\frac{3\,810\,179}{3\,879\,263}$$

Por lo tanto, la estimación de la población de San José, al 30 de junio del 2000 (fecha del censo), viene dada por:

$$\frac{3\,810\,179}{3\,879\,263} \cdot 282\,364 \approx 277\,336$$

Entonces el error porcentual para la estimación anterior es:

$$EP = \frac{(277\,336 - 309\,672)}{309\,672} \cdot 100 \approx -10,4\%$$

Se puede decir que el método de Razón Censal produjo una subestimación de un 10,4% con respecto a su valor censal de la población de San José.

**Método de Diferencia de Tasas:** Este método es una variante del procedimiento de Razón Censal. La única modificación radica en la forma de estimar la tasa de ocurrencia del indicador sintomático. En el ejemplo anterior se obtuvieron los valores  $r(\text{San José})_{1984} = 0,74229$ ,  $r(\text{Costa Rica})_{1984} = 0,56863$  y  $r(\text{Costa Rica})_{2000} = 0,56772$

Por medio del procedimiento Tasa Correlación se supone que la tasa de electores de San José, para 1984, viene dada por:

$$r(\text{San José})_{2000} = r(\text{San José})_{1984} + [r(\text{Costa Rica})_{2000} - r(\text{Costa Rica})_{1984}]$$

Con lo cual, según esta técnica la estimación de  $r(\text{San José})_{2000}$  es:

$$\hat{r}(\text{San José})_{2000} = 0,74229 + [0,56772 - 0,56863] = 0,74138$$

Así, la estimación simple de la población para 1984 está dada por:

$$\hat{P}(\text{San José})_{2000} = \frac{\text{Elec. Insc.}_{2000}}{\hat{r}(\text{San José})_{2000}} = \frac{209\,260}{0,74138} \approx 282\,257$$

Para este caso, la suma de las estimaciones simples para todos los cantones es 3 873 405 y la estimación de la población de San José es:

$$\frac{3\ 810\ 179}{3\ 873\ 405} \cdot 282\ 257 \approx 277\ 650$$

El error porcentual asociado es :

$$EP = \frac{(277\ 650 - 309\ 672)}{309\ 672} \cdot 100 \approx -10,3\%$$

Por lo que se concluye que la estimación poblacional subestimó el valor censal en 10,3%.

**Método de Razón Correlación:** Para la aplicación de este método se requiere información de dos censos anteriores al momento de la estimación:

La técnica utiliza el modelo de regresión lineal:

$$Y_u = a_0 + a_1 X_u$$

donde  $a_i$  ( $i = 0,1$ ) corresponde a los coeficientes obtenidos por mínimos cuadrados, y para este caso:

$$Y_u = \frac{\left( \frac{P_{84}(u)}{F_{84}(M)} \right)}{\left( \frac{P_{73}(u)}{F_{73}(M)} \right)} \quad X_u = \frac{\left( \frac{S_{84}(u)}{S_{84}(M)} \right)}{\left( \frac{S_{73}(u)}{S_{73}(M)} \right)}$$

para todo cantón  $u$ .

$P(u)$  es la población censal correspondiente al cantón  $u$

$P(M)$  es la población censal de Costa Rica

$S(u)$  es el número de electores inscritos en el padrón electoral para el cantón  $u$

$S(M)$  es el número de electores inscritos en el padrón electoral para Costa Rica

De este modo,  $Y_u$  y  $X_u$  representan las razones de cambio de las proporciones que representa cada cantón del total del país, para el período 73-84. Para San José, estas razones están dadas por:

$$Y_{\text{San José}} = \frac{\left( \frac{241\,464}{2\,416\,809} \right)}{\left( \frac{215\,441}{1\,871\,790} \right)} = 0,86803 \quad X_{\text{San José}} = \frac{\left( \frac{179\,237}{1\,374\,278} \right)}{\left( \frac{120\,647}{791\,033} \right)} = 0,85512$$

Una vez determinadas las razones de cambio en el período 73-84 para los 81 cantones, se procede a utilizar mínimos cuadrados para obtener los coeficientes del modelo. El siguiente cuadro presenta los resultados dados por el paquete estadístico SPSS:

Valor estadístico	Estimador	Significancia
<b>ANOVA</b>	F = 310,7 (1,79)gl	0,000
<b>R<sup>2</sup></b>	0,80	
<b>Constante</b>	0,261	0,000
<b>Coefficiente</b>	0,730	0,000

La prueba “F” indica que el modelo es altamente significativo. El valor de  $R^2 = 0,80$ , muestra que el 80% de la variación de la razón de cambio en la proporción poblacional es explicada por la variación de la razón de cambio en la proporción del número de electores inscritos.

El modelo de regresión lineal estimado viene dado por:

$$\hat{Y}_u = 0,261 + 0,730 \cdot X_u$$

La razón de cambio en la proporción poblacional para San José, en el período 1984-2000, se estima por:

$$\hat{Y}_{\text{San José}} = 0,261 + 0,730 \cdot X_{\text{San José}}$$

Como para el período 1984-2000 la razón de cambio en la proporción del padrón es:

$$X_{\text{San José}} = \frac{\left( \frac{209\,260}{2\,163\,130} \right)}{\left( \frac{179\,237}{1\,374\,278} \right)} = 0,74174$$

entonces el valor estimado para la razón de cambio en la proporción de la población sería:

$$\hat{Y}_{\text{San José}} = 0,261 + 0,730 \cdot 0,74174 = 0,80247$$

Para el período 1984-2000 se tiene que:

$$\hat{Y}_{\text{San José}} = \frac{\hat{P}(\text{San José})_{2000}}{\left( \frac{241\,464}{2\,416\,809} \right)} = \frac{3\,810\,179}{\left( \frac{241\,464}{2\,416\,809} \right)}$$

Con lo que se tiene que:

$$\hat{Y}_{\text{San José}} = \frac{\hat{P}(\text{San José})_{2000}}{\left( \frac{241\,464}{2\,416\,809} \right)} = 0,80247$$

Entonces, la estimación simple para la población de San José al 30 de junio del 2000 (fecha del censo), viene dada por :

$$\hat{P}(\text{San José})_{2000} = 0,80247 \cdot \left( \frac{241\,464}{2\,416\,809} \right) \cdot 3\,810\,179 = 305\,481$$

No obstante, la suma de las estimaciones poblacionales para los 68 cantones es 2434460, que no reproduce la población total de Costa Rica (2416809). Por esta razón, es necesario prorratear las estimaciones simples obtenidas, por lo que la estimación poblacional para San José es:

$$\frac{3\,810\,179}{3\,837\,348} \cdot 305\,481 \cong 303\,318$$

Debido a que el valor censal de la población de San José fue 309 672, el error porcentual de la estimación es:

$$EP = \frac{(303\,318 - 309\,672)}{309\,672} \cdot 100 \cong -2,1\%$$

Lo cual quiere decir, que la estimación realizada por este método subestimó en un 2,4% la población censal de San José.

**Método de Tasa Correlación:** En la aplicación del método de Tasa Correlación se mantienen los principios básicos utilizados por el Método de Razón Correlación. La principal diferencia entre los dos métodos es la forma en que se aplica la razón de cambio en las proporciones. Formalmente el modelo se puede expresar por:

$$Y_u^* = a_0 + a_1 X_u^*$$

donde :

$$v_u^* = \frac{\ln(Y_u^*)}{k} \quad X_u^* = \frac{\ln(X_u)}{k}$$

$Y_u$  y  $X_u$  están definidos tal y como se indicó para la técnica de Razón Correlación y  $k$  corresponde al intervalo de tiempo transcurrido entre los censos.

De acuerdo con los cálculos hechos anteriormente, para San José, en el período 73-84,  $Y_{\text{San José}} = 0,86803$  y  $X_{\text{San José}} = 0,85512$ . Además  $k = 11,06$  y representa el tiempo transcurrido entre los censos del 73 y 84 en años. Entonces:

$$\hat{Y}_{\text{San José}}^* = \frac{\ln(0,86803)}{11,06} = -0,01280 \quad X_{\text{San José}}^* = \frac{\ln(0,85512)}{11,06} = -0,01415$$

Después de calculados estos valores en el período 73-84 para los demás cantones, se utilizan de nuevo mínimos cuadrados para obtener los coeficientes  $a_0$  y  $a_1$ . El siguiente cuadro presenta los resultados obtenidos por medio del paquete estadístico SPSS, respecto a la aplicación de dicho método.

Valor estadístico	Estimador	Significancia
ANOVA	F = 384,3 (1,79)gl	0,000
R <sup>2</sup>	0,83	
Constante	-0,001	0,346
Coefficiente	0,791	0,000

Nuevamente la prueba “F” muestra que el modelo es altamente significativo. El valor de  $R^2 = 0,83$ , indica que el 83% de la variación de  $Y_u^*$  es aclarada por medio de la variación en  $X_u^*$ . Puede notarse que el porcentaje de explicación de este modelo es ligeramente superior al modelo anterior.

El modelo de regresión lineal estimado es:

$$\hat{Y}_u^* = -0,001 + 0,791 \cdot X_u^*$$

Entonces para el cantón Central de San José, la estimación viene dada por:

$$\hat{Y}_{\text{San José}}^* = -0,001 + 0,791 \cdot X_{\text{San José}}^*$$

Para el período intercensal 1984-2000, el tiempo transcurrido es  $k = 16,07$ , entonces:

$$X_{\text{San José}}^* = \frac{\ln \left[ \frac{\left( \frac{209\ 260}{2163130} \right)}{\left( \frac{179\ 237}{1374278} \right)} \right]}{16,07} = -0,01859$$

con lo cual:

$$\hat{Y}_{San\ José}^* = -0,001 + 0,791 \cdot (-0,01859) = -0,01570$$

Entonces

$$\hat{Y}_{San\ José}^* = \frac{\ln \left( \frac{P(San\ José)_{2000}}{3\ 810\ 179} \right)}{\ln \left( \frac{241\ 464}{2\ 416\ 809} \right)} = -0,01570$$

Por lo tanto, la estimación simple para la población de San José viene dada por :

$$P(San\ José)_{2030} = \left( \frac{241\ 464}{2\ 416\ 809} \right) \cdot 3\ 810\ 179 \cdot e^{(-0,01570 \cdot 16,07)} = 295\ 790$$

La suma de las estimaciones poblacionales para los 81 cantones es 3 803 412. Por lo tanto, la estimación poblacional para San José es:

$$\frac{3\ 810\ 179}{3\ 803\ 412} \cdot 295\ 790 = 296\ 316$$

El error porcentual en la estimación es:

$$EP = \frac{(296\ 316 - 309\ 672)}{309\ 672} \cdot 100 \cong -4,3\%$$

Lo que da una subestimación del valor censal de un 4,3%.

Las cuatro estimaciones realizadas utilizan un único indicador sintomático, pero la experiencia indica que las mejores estimaciones se producen cuando se combinan más de un indicador. El siguiente cuadro muestra un análisis de los errores que se produjeron al utilizar estas técnicas para estimar los cantones de Costa Rica en los últimos dos años censales.

Indicadores sintomáticos utilizados	Razón Censal		Diferencia de Tasas		Razón Correlación		Tasa Correlación	
	1984	2000	1984	2000	1984	2000	1984	2000
<b>Padrón electoral y nacimientos</b>								
Error porcentual promedio	4,7	4,0	4,7	4,2	4,1	4,3	4,0	4,6
Error porcentual máximo	21,6	11,7	20,7	20	12,1	13,8	12,5	14,7
Porcentaje de errores menores a 5%	66,2	69,1	64,7	59,3	72,1	63,0	72,1	63,0
Porcentaje de errores mayores 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>								
Error promedio (valores absolutos)	4,8	3,4	5,4	4,5	3,8	3,7	3,8	3,7
Error máximo	17,5	15,5	17,2	19,2	13,5	9,9	14,2	9,5
Porcentaje de errores menores a 5%	55,9	80,2	57,4	67,9	72,1	71,6	72,1	70,4
Porcentaje de errores mayores a 10%	8,8	3,7	14,7	12,3	2,9	0,0	1,5	0,0

**Nota:** todo el análisis se hizo con el valor absoluto de errores porcentuales.

Fuentes: Chaves, Edwin (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.

Chaves, Edwin. **Validación de los indicadores sintomáticos en las estimaciones poblacionales: Censo del 2000.** Jornadas de Análisis Estadístico. Escuela de Estadística, UCR. 2002.

Puede notarse como la combinación de los indicadores: padrón electoral, nacimientos y matrícula escolar reducen los errores en las estimaciones de población.