

EL MUESTREO DE AREAS: UNA ALTERNATIVA PARA ESTIMAR CARACTERISTICAS POBLACIONALES ANTE LA AUSENCIA DE DATOS CENSALES

Edwin A. Chaves Esquivel

Escuela de Estadística, UCR
Escuela de Matemática, UNA
E-mail: echaves@rad.fod.ac.cr

RESUMEN

Este artículo discute el uso del muestreo de áreas en la obtención de estimaciones de características poblacionales de una determinada región. Para ello se utilizan los segmentos censales del último censo de población y vivienda. La técnica fue aplicada en el cantón de San Rafael de Heredia y los resultados demuestran que el método es una buena alternativa, para la estimación de ciertas características poblacionales ante la carencia de datos censales.

ABSTRACT

This paper discusses the use of area sampling for providing estimates of population characteristics of a determined area. For this purpose census tracts of the last Population and Housing Census are used. The technique was applied to data gathered in the canton of San Rafael of Heredia. The results show that the technique is a good alternative for the estimation of certain population characteristics when census data ARE not available.

INTRODUCCION

La incorporación de la variable población en la planificación del desarrollo, convirtió a las estimaciones y proyecciones poblacionales, referidas a áreas de diversos tamaños, en un recurso

demográfico necesario (CHAVES 1997). La estimación de las principales características poblacionales (tamaño, número de viviendas, estructura por edad y sexo, población económicamente activa, nivel de escolaridad, etc.) constituye el elemento básico para la planeación y preparación de los programas de desarrollo, para la formulación de las políticas de acción gubernamental y privada, así como para la ejecución y evaluación de dichos planes (RINCON 1990).

De este modo, las estimaciones de las principales características poblacionales de una localidad constituyen un instrumento fundamental para el diagnóstico de sus necesidades y para estimar la demanda en la formulación de proyectos. Por ejemplo, el número de habitantes y su composición por sexo y edad debe considerarse al crear nuevos centros educativos, clínicas de atención médica, campos deportivos, puestos de vigilancia o para implementar servicios generales, como electricidad, agua, teléfono, alcantarillado, recolección de basura, etc. Estos elementos revisten especial interés en el proceso de democratización de un país (CELADE, MIDEPLAN y DGEC 1990).

El sector privado también requiere de estas estimaciones para establecer sus programas y políticas empresariales de desarrollo futuro. Ellas se convierten en la fuente de información básica de clientes y consumidores potenciales y, a la vez, son el fundamento de sus demandas de mano de obra.

Por ejemplo, para determinar las áreas de atracción al construirse un nuevo centro comercial, es necesario, entre otros aspectos, contar con estimaciones confiables del número de habitantes y de sus principales características, con el fin de buscar un balance apropiado entre inversión y número de consumidores potenciales.

Estimaciones eficientes en las características poblacionales en áreas menores se tienen fundamentalmente para los años censales. Sin embargo, durante el período intercensal, el cual con frecuencia supera los diez años, quienes deben tomar decisiones con base en estos datos, requieren de información precisa, la que no se posee en general.

En esta materia es poco lo que se ha hecho en Costa Rica. Se han efectuado intentos por estimar el tamaño poblacional de una determinada área, éstos suponen que las condiciones del pasado se mantienen en el tiempo, llegándose a estimaciones y proyecciones poblacionales poco confiables. El método más común, y que se utiliza desde hace muchos años, consiste en suponer un crecimiento natural de la población. Es decir, se suman a la población censal los nacimientos y se restan las defunciones ocurridas en el período de interés. No obstante, al no considerar el movimiento migratorio, las estimaciones son cada vez menos confiables, conforme se alejan del año censal. Esto provoca una subestimación en aquellas zonas de atracción poblacional y una sobrestimación en las de expulsión (CHAVES 1997). La principal dificultad en considerar los movimientos migratorios, dentro del proceso de estimación, consiste en la carencia de registros, con el detalle y la calidad necesarios. Pese a esta problemática, estas estimaciones son las más utilizadas por parte de estudiantes, profesores, investigadores o por instituciones públicas y privadas.

El problema se agrava en la actualidad, porque trece años después del último censo de población y vivienda, no se ha decidido en qué año será el próximo y posiblemente en lo que queda del siglo no se cuente con esta información.

Este documento ofrece una contribución en este sentido. Por medio de la aplicación del mues-

treo de áreas (KISH 1972), se presenta una alternativa para estimar ciertas características poblacionales de una localidad, como tamaño poblacional, número de viviendas, estructura por edad, sexo, y otras, que no fueron consideradas en el estudio. La técnica se aplica en el cantón de San Rafael de Heredia y consiste en utilizar el muestreo probabilístico para seleccionar una muestra de segmentos censales de este cantón (COCHRAN 1987), de manera que comparando su comportamiento en el último censo (1984) con el comportamiento actual, sea posible estimar los cambios ocurridos por medio de ciertos cálculos matemáticos.

MATERIALES Y METODOS

Aunque se han sugerido algunas formas indirectas de estimar la población en áreas administrativas menores (CHAVES 1997), las mejores aproximaciones se obtienen por medio de la observación directa. En seguida se describen los principales fundamentos metodológicos de una de estas técnicas, que se aplica posteriormente en la determinación de algunas características poblacionales del cantón de San Rafael de Heredia.

Una de las principales ventajas de este método es que, aparte de producir cálculos totales de población, permite estimar ciertas particularidades para diferentes grupos, tales como sexo, edad, ocupación, analfabetismo, etc. Otra ventaja radica en que el área de estimación no se circunscribe a límites político-administrativos, lo que posibilita obtener estimaciones poblacionales a otros niveles geográficos, fundamentalmente útiles para el comercio.

El esquema de la aplicación del muestreo de áreas en el proceso de estimación de la población total, involucra cinco etapas (RIVES 1982):

1. En la primera se requiere dividir el área geográfica en un número grande "N" de unidades menores. Estas unidades menores constituyen grupos de viviendas y corresponden a las llamadas "unidades primarias de muestreo" (COCHRAN 1987). Dependiendo del área en consideración es posible tomar manzanas o segmentos censales, para los cuales existe cartografía, de manera que

simplifique el trabajo. Por las características de Costa Rica, para esta etapa lo ideal es trabajar con los segmentos censales.

2. La segunda etapa requiere seleccionar una muestra de "n" grupos del total del área (N). La selección debe efectuarse con base en los grupos establecidos en el momento del último censo, puesto que el modelo base involucra el cambio entre los datos censales y la información obtenida para el período poscensal.
3. La tercera etapa del proceso de estimación requiere determinar la información censal sobre el número de viviendas y de personas que residían en dicho grupo.
4. La cuarta etapa consiste en el denominado "trabajo de campo" (GOMEZ 1985). Se debe recorrer cada uno de los grupos o segmentos seleccionados y recoger la información del número de viviendas. Además, cada vivienda debe proveer información sobre la cantidad de personas que usualmente residen en ella, siguiendo la misma definición censal de residencia habitual. Si se requiere información para grupos particulares, entonces, se debe solicitar la información específica para cada miembro de la vivienda.
5. La etapa final del procedimiento consiste en efectuar los cálculos. Sea y_i la población correspondiente al i-ésimo segmento muestreado y sea x_i la población del mismo segmento obtenida en el censo, entonces:

$$Y = \sum_{i=1}^n y_i \quad \text{y} \quad X = \sum_{i=1}^n x_i \quad (1)$$

corresponden a los totales de la población para los segmentos muestreados en el momento de la observación y en el momento censal, respectivamente.

Si P_e corresponde a la población del área en consideración para el último censo y P_c a la población estimada por medio de la muestra, entonces se tiene que:

$$P_e = \frac{Y}{X} P_c \quad (2)$$

El valor Y/X corresponde al estimador de razón del crecimiento poblacional del período poscensal.

La población estimada P_e está sujeta a errores muestrales y no muestrales del estimador de razón y de los errores de cobertura del censo en el cálculo de P_c . Los errores de muestreo pueden ser controlados efectivamente por medio del diseño de una muestra eficiente y los no muestrales a través del control de la calidad del procedimiento. Los errores de cobertura censal traen mayores problemas, pues no es posible saber cuál fue el error cometido en cada área.

Cuando el número de grupos o segmentos de viviendas es grande, generalmente mayor que treinta, y éstos producen coeficientes de variación para los valores promedio de los x_i y de los y_i , relativamente pequeños, en general menos del 10%, entonces, el estimador de razón Y/X se distribuye aproximadamente normal en muestras repetitivas (COCHRAN 1987). En estas condiciones, la variancia para muestras del estimador P_e está dada por la expresión:

$$\frac{N^2(1-f)}{n(N-1)} \sum_{i=1}^n (y_i - Rx_i)^2 \quad (3)$$

donde R es la razón verdadera de crecimiento poblacional para el período poscensal y f es la fracción de muestreo (n/N). La ecuación anterior supone un muestreo aleatorio simple. El estimador para la variancia poblacional de P_e puede expresarse como:

$$\frac{N^2(1-f)}{n(n-1)} \sum_{i=1}^n (y_i - rx_i)^2 \quad (4)$$

Donde r es la razón de crecimiento muestral para el período poscensal. La última ecuación es una expresión aproximada, la cual presenta un sesgo del orden $1/n$ (COCHRAN 1987).

Si se consideran características particulares de la población o de las viviendas, el estimador de la proporción de unidades pertenecientes a una de esas categorías viene dado por:

$$\hat{p} = \frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n Y_i} \quad (5)$$

donde Y_i es el número de elementos en el i -ésimo segmento y a_i corresponde al número de elementos pertenecientes a la categoría de interés en este segmento.

Todo lo expuesto está dirigido a la población residente en viviendas simples. Si en el área de interés se encuentran personas que residen en residencias colectivas, tales como prisiones, residencias estudiantiles, conventos, etc., deben ser consideradas por separado y agregadas a la estimación efectuada.

Dependiendo de las condiciones del área seleccionada, se pueden introducir algunas variaciones al método, como estratificar por áreas, por crecimiento residencial en los últimos años, por características urbano-rurales, etc., con el objetivo de evitar sesgos en las estimaciones.

RESULTADOS

Se decidió seleccionar un cantón que cumpliera ciertas características de urbanidad, migración y tamaño, de modo que sus resultados permitieran valorar su factibilidad. Después de analizar varias posibilidades se escogió el cantón de San Rafael de Heredia, considerando que cumplía los requisitos mínimos aceptables. En San Rafael, para 1984, el 68% de su población vivía en zona rural. Su tasa de emigración para el período 1979-1984 fue de 20% y la de inmigración de 28%, lo que indica un aceptable movimiento migratorio. Su densidad poblacional se ha estimado que cambiará de 455 h/km² en 1984 a 786 h/km² en el 2000, mientras que su tasa de crecimiento se ha proyectado para el período 1985-2000 en 2,9% (CELADE, MIDEPLAN y DGEC 1990).

Para la aplicación de la técnica se utilizó el marco muestral de 84 segmentos censales que poseía San Rafael para 1984. Sin embargo, según información suministrada por personeros de la Municipalidad del cantón, en tres de estos segmentos se ha dado un crecimiento poblacional mucho más elevado que en los otros, debido a la creación de proyectos urbanísticos. Por tal razón, se tomó la decisión de excluirlos del marco muestral y considerarlos por separado. De no dárseles un trato independiente a estos segmentos y de haber sido seleccionado al menos uno en la muestra, se presentaría una sobrestimación de la población. Por otro lado, de no seleccionar alguno, se obtendría entonces una subestimación. Con los 81 segmentos restantes se seleccionó una muestra simple al azar de 25 (KISH 1972), considerándose que con ellos se lograban los objetivos del estudio.

Cada segmento muestreado se recorrió completamente, actualizando el número de viviendas; en cada vivienda se preguntó por el número de personas que allí residía. Además, para cada segmento, en forma sistemática, se seleccionó una de cada cinco viviendas, con el fin de preguntar el sexo y la edad de sus residentes. Este mismo procedimiento se llevó a cabo en los tres segmentos que fueron separados del marco, con el objetivo de agregarlos a los resultados de la muestra.

El trabajo de campo se llevó a cabo entre el 18 de octubre y el 9 de noviembre de 1996. Los resultados se muestran en el cuadro 1.

Se presenta un fuerte crecimiento en casi todos los segmentos, con la excepción de algunos que por sus características de densidad o urbanidad no sufrieron grandes cambios. También se corrobora el fuerte crecimiento que se dio en los tres segmentos tratados por separado, lo cual justifica la decisión de separarlos. Los 25 segmentos muestreados produjeron 2116 viviendas con un total de 9315 personas, para un promedio de 4,40 personas por vivienda. Del mismo modo, en los tres segmentos separados se contaron 901 viviendas con 3974 personas, para un promedio de 4,41 personas por vivienda.

Cuadro 1
Población total y número de viviendas por segmento censal muestreado.
San Rafael de Heredia, 1996

<i>Muestra</i> Segmento	<i>Viviendas</i>		<i>Población</i>	
	<i>Censo 1984</i>	<i>Conteo 1996</i>	<i>Censo 1984</i>	<i>Conteo 1996</i>
1-03	61	85	303	426
1-05	68	104	319	451
1-07	53	55	226	242
1-09	49	53	201	200
1-13	69	89	301	433
1-16	65	82	314	387
1-18	48	43	242	163
1-21	86	137	375	560
2-01	69	76	326	302
2-05	60	78	290	319
2-09	70	64	355	311
2-19	64	79	298	313
2-20	71	92	358	453
3-01	60	70	261	269
3-05	65	92	343	374
3-06	65	85	274	338
3-07	65	138	207	667
3-10	55	100	262	486
4-01	61	101	259	489
4-05	32	77	129	361
4-11	70	117	320	463
4-16	34	52	138	223
5-01	75	112	351	479
5-03	35	62	119	259
5-04	55	73	241	347
Total	1505	2116	6812	9315
Segmentos separados				
1-25	85	199	381	849
2-15	54	319	271	1474
3-08	36	383	207	1651
Total	175	901	859	3974

A continuación se detallan los cálculos realizados para la obtención de las estimaciones. Sean:

P_c : Población según censo de 1984

P_e : Población estimada

V_c : Número de viviendas según censo de 1984

V_e : Número estimado de viviendas

r : Razón de crecimiento poblacional

r_i : Razón de crecimiento en viviendas

y_i : Población actual del i -ésimo segmento

x_i : Población censal del i -ésimo segmento

a_i : Número de personas muestreadas en el i -ésimo segmento con característica de interés

Y_i : Número de personas muestreadas en el i -ésimo segmento

$$r = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i} = \frac{9315}{6812} = 1,3674$$

El estimador de la razón para el crecimiento total de la población de los 81 segmentos considerados en el marco, viene dado por:

La aplicación de la razón de crecimiento estimada a la población censal de 1984 (21212), da la estimación poblacional siguiente para los 81 segmentos:

$$P_e = r \cdot P_c = \frac{9315}{6812} \cdot 21212 = 29006$$

Si se agrega la población de los tres segmentos separados, se obtiene que la estimación poblacional para San Rafael de Heredia a octubre de 1996 es:

$$P = 29006 + 3974 = 32980$$

El coeficiente de variación para el promedio de la población censal de los 25 segmentos es 5,3%, mientras que para el promedio poblacional de la muestra es 7,3%. Ambos valores son menores al 10%, y si se supone que la muestra de 25 segmentos es grande, de manera que pueda considerarse el estimador poblacional utilizado como normalmente distribuido (COCHRAN 1987), se tiene que la variancia estimada de P_e está dada por:

$$\begin{aligned} Var(P_e) &= \frac{N^2(1-f)}{n(n-1)} \sum_{i=1}^n (y_i - rx_i)^2 = \\ &= \frac{25}{81} \frac{355998,31}{25 \cdot 24} = 2691347,22 \end{aligned}$$

Con error estándar igual a:

$$ee(P_e) = \sqrt{2691347,22} = 1641$$

Los límites del intervalo de confianza del 90% para la estimación de la población de San Rafael de Heredia a octubre de 1996 son:

$$\text{Límite inferior: } P - t_{n-1, \alpha/2} \cdot ee(P_e) = 32980 - 1,71 \cdot 1641 = 30174$$

$$\text{Límite superior: } P + t_{n-1, \alpha/2} \cdot ee(P_e) = 32980 + 1,71 \cdot 1641 = 35786$$

Por otro lado, la estimación de la razón de cambio en viviendas para los 25 segmentos muestreados es:

$$r_l = \frac{\sum_{i=1}^n t_i}{\sum_{i=1}^n w_i} = \frac{2116}{1505} = 1,4060$$

Y así, la estimación del número de viviendas de los 81 segmentos correspondientes a la población considerada es:

$$V_e = r_l \cdot V_c = \frac{2116}{1505} \cdot 4780 = 6721$$

Al agregar a este valor las 901 viviendas existentes en los tres segmentos separados se obtiene que, para octubre de 1986, el número total estimado de viviendas para San Rafael de Heredia es:

$$V = 6721 + 901 = 7622$$

La estimación de la variancia de V_e es:

$$\begin{aligned} Var(V_e) &= \frac{N^2(1-f)}{n(n-1)} \sum_{i=1}^n (t_i - r_l w_i)^2 = \\ &= \frac{25}{81} \frac{8342,46}{25 \cdot 24} = 63069,0 \end{aligned}$$

y el correspondiente error estándar es:

$$ee(V_e) = \sqrt{63069,0} = 251$$

De este modo, los límites del intervalo de confianza del 90% para la estimación de las viviendas de San Rafael de Heredia a octubre de 1996 son:

$$\text{Límite inferior: } V - t_{n-1, \alpha/2} \cdot ee(V_e) = 7622 - 1,71 \cdot 251 = 7193$$

$$\text{Límite superior: } V + t_{n-1, \alpha/2} \cdot ee(V_e) = 7622 + 1,71 \cdot 251 = 8051$$

Con respecto al estudio de las características sexo y edad de la población, se contó con una muestra de sólo 380 viviendas con un total de 1706 personas. En este caso, el porcentaje de "no respuesta" fue aproximadamente un 10%, correspondiente a personas que se negaron a responder o no fue posible localizarlas en sus viviendas. Del total de personas muestreadas, 837 fueron hombres, por lo que la proporción estimada de varones para los 81 segmentos es:

$$\hat{p} = \frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n Y_i} = \frac{837}{1706} = 0,491$$

Por otro lado, para los tres segmentos considerados separadamente se observaron 157 viviendas con 796 personas, de las cuales 408 fueron hombres, para una proporción de varones estimada de:

$$\hat{p} = \frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n Y_i} = \frac{408}{796} = 0,513$$

De este modo la proporción de varones estimada para el cantón de San Rafael de Heredia es:

$$\frac{0,491 \cdot 25 + 0,513 \cdot 3}{28} = 0,493$$

El cuadro 2 muestra un resumen de los resultados en los cálculos de las proporciones por

edad y sexo para la población de San Rafael, usando una estrategia similar a la descrita anteriormente para estimar la proporción de varones.

DISCUSION

El análisis mostró que la técnica es una buena alternativa para la obtención no sólo de la información sobre población total, sino de otras características poblacionales, como número de viviendas, promedio de personas por vivienda, razón de crecimiento poblacional y de viviendas para el período de interés, estructura por edad y sexo, y otras, que podrían considerarse en el futuro. Este procedimiento permite determinar errores muestrales, permitiendo efectuar inferencias estadísticas para las estimaciones obtenidas. Otra de las ventajas radica en que la técnica hace posible darle un trato separado a aquellos segmentos de mayor crecimiento, con lo que se logra una mayor precisión en las estimaciones producidas.

Los principales problemas que acarrea la aplicación de la técnica se fundamentan en el tiempo invertido en el estudio y el costo de su realización. Cada una de las etapas que compone el estudio requiere de una inversión grande de tiempo y recursos. Debido a estas consideraciones, no se justificaría su aplicación si lo que se busca es únicamente estimar el tamaño poblacional de un área administrativa, ya que se podría obtener estimaciones de una calidad similar por medio de otros procedimientos (CHAVES 1997).

Sin embargo, la estrategia utilizada cobra especial atractivo ante la problemática del país en materia de población, donde trece años después del último censo se desconoce la fecha de realización del próximo. Este método podría ser aplicado para la determinación de las características poblacionales en aquellas áreas que sean de interés para investigadores tanto pertenecientes al sector público como al privado.

Finalmente, debe recordarse que la exactitud de un método de estimación no puede ser generalizada, y que por más robusto que pueda ser un determinado procedimiento, no se puede asegurar que su aplicación a diferentes situaciones conduzca a resultados que reflejen la realidad. Al

Cuadro 2
Distribución proporcional de la población de San Rafael de Heredia
por sexo y grupos de edad

Grupos de edad	Muestra de 25 segmentos		Tres segmentos separados		Total
	$\sum_{i=1}^n a_i$	Proporción estimada p_1	$\sum_{i=1}^n a_i$	Proporción estimada p_2	Proporción estimada $\frac{p_1 \cdot 25 + p_2 \cdot 3}{28}$
Hombres					
0-9	157	0,092	92	0,115	0,094
10-19	179	0,105	102	0,128	0,107
20-29	145	0,085	79	0,099	0,087
30-39	141	0,083	64	0,080	0,083
40-49	100	0,059	45	0,056	0,059
50-59	46	0,027	13	0,016	0,026
60 y más	69	0,040	13	0,016	0,037
Total	836	0,491	408	0,513	0,493
Mujeres					
0-9	146	0,086	98	0,123	0,090
10-19	203	0,119	93	0,117	0,119
20-29	132	0,077	57	0,072	0,076
30-39	153	0,090	83	0,104	0,092
40-49	84	0,049	19	0,024	0,046
50-59	62	0,036	19	0,024	0,035
60 y más	89	0,052	19	0,024	0,049
Total	870	0,509	389	0,487	0,507
Ambos sexos					
0-9	303	0,178	190	0,238	0,184
10-19	381	0,223	196	0,246	0,225
20-29	278	0,163	137	0,172	0,164
30-39	294	0,172	147	0,184	0,173
40-49	184	0,108	64	0,080	0,105
50-59	108	0,063	32	0,040	0,061
60 y más	158	0,093	32	0,040	0,087
Total	1706	1,000	797	1,000	1,000

analizar las estimaciones deben considerarse las hipótesis correspondientes a los métodos y la calidad de la información que se maneja. Es importante, entonces, tener un juicio crítico y no caer en el uso mecánico de los métodos desarrollados, ni de los resultados obtenidos.

REFERENCIAS

- Anónimo. 1990. Estimaciones y proyecciones de población cantonal por sexo y grupos de edades: 1975-2000. MIDEPLAN, CELADE y DGEC. San José, Costa Rica.
- 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.
- Cochran, William G. 1987. Técnicas de muestreo. CECSA. México D.F.
- DGEC. Censos de Población y Vivienda de 1973, 1984. San José, Costa Rica, 1974 y 1987.
- Gómez, Miguel. 1985. Elementos de estadística descriptiva. EUNED. San José, Costa Rica.
- Kish, Leslie. 1972. Muestreo de encuestas. John Wiley & Sons. New York, EUA.
- Lee, Everett S. y Harold F. Goldsmith. 1982. Population estimates: methods for small area analysis. SAGE Publications. Beverly Hills, California.
- Malec, D. y J. Katzoff. 1983. Application of some common spatial models to population data. En: Reunión Anual de la Asociación Estadísticos Americanos, Toronto, Ontario.
- Rincón, Manuel. 1990. Teoría y métodos para la preparación de estimaciones y proyecciones de población. CELADE. San José, Costa Rica.
- Rives, Norfleet W. 1982. Assessment of a survey approach. En: Lee, Everett S. y Harold F. Goldsmith. Population estimates: methods for small area analysis (pp. 79-96). SAGE Publications. Beverly Hills, California.