

**USO DE VARIABLES SINTOMÁTICAS PARA ESTIMAR
LA DISTRIBUCIÓN ESPACIAL DE POBLACIÓN**
Aplicación a los municipios de Rio Grande do Sul, Brasil ¹

Maria de Lourdes Teixeira Jardim
Fundación de Economía y Estadística
Rio Grande do Sul - Brasil

RESUMEN

El uso de variables sintomáticas para estimar el crecimiento regional de áreas menores aparece muy tímidamente en Brasil como indicador de cambios de tendencia del crecimiento de la población.

En este trabajo se presenta algunos métodos para la actualización de información de población para los municipios del Estado de Rio Grande do Sul en Brasil. En este sentido, se propone la implantación de metodología de estimación de población por medio de variables sintomáticas.

Diversos autores han utilizado metodologías basadas en el uso de variables socioeconómicas como indicadores del crecimiento demográfico. En algunas se estima la población urbana del Estado, en otras se aplica una desagregación a nivel de municipios. En otros cinco trabajos se aprecia que las estimaciones de población de áreas menores utilizaron variables sintomáticas, entre las cuales cabe citar la media de personas ocupadas, el número de establecimientos industriales, datos de

¹ Este documento es un extracto de la tesis presentada al Centro de Desarrollo y Planificación Regional de la Facultad de Ciencias Económicas de la Universidad de Minas Gerais, para optar al título de Master en Demografía.

los censos agropecuarios, número de camas de hospitales, cifras de electores, tiraje de periódicos y matrícula en la educación primaria.

El objetivo de este trabajo consiste en evaluar la eficacia de diversas variables sintomáticas para distribuir la población total del estado en sus municipios a lo largo del decenio de 1990, a partir de los resultados del censo demográfico de 1991.

THE USE OF SYMPTOMATIC VARIABLES IN ESTIMATING SPATIAL DISTRIBUTION OF THE POPULATION IN THE MUNICIPALITIES OF RIO GRANDE DO SUL, BRAZIL

ABSTRACT

The use of symptomatic variables to estimate the regional growth of small areas is just beginning to be applied in Brazil as an indicator of changes in population growth trends.

This article presents a number of methods for updating the population data for the municipalities of the State of Rio Grande do Sul in Brazil and proposes the implementation of a population estimation methodology based on the use of symptomatic variables.

Several authors have used methodologies based on socio-economic variables as indicators of demographic growth. Some of these methodologies are used to estimate the urban population of the State, while others allow for disaggregation at the municipal level. Another five studies have used symptomatic variables for estimating the population of small areas, including average number of employed, number of industrial establishments, data from agricultural surveys, number of hospital beds, electoral statistics, newspaper circulation figures and statistics on enrolment in primary education.

The objective of this study is to evaluate the effectiveness of various symptomatic variables for estimating the State's total population distribution by municipalities throughout the 1990s based on the results of the 1991 census.

INTRODUCCIÓN

Las estimaciones de población son los datos más solicitados por planificadores e investigadores en general, debido a que la población de una área geográfica constituye una base tanto para la planificación del sector público como para proyectos de empresas privadas; en ambos casos se necesita conocer, entre otros, la demanda potencial de productos, mano de obra y servicios. En un proceso dinámico, la magnitud de la población influye en la situación económica y social del área en consideración y es influenciada por ésta.

A pesar de su importancia, la magnitud de la población es un dato muy difícil de estimar cuando se trata de áreas menores. Esto se debe a la dinámica de los cambios que afectan a las áreas menores y al gran volumen de información que debe manejarse, lo que impide prever cambios bruscos en la tendencia del crecimiento de estas áreas en períodos intercensales sin un análisis específico de la realidad socioeconómica de cada una de ellas.

Además de la magnitud total de la población, también son importantes, sobre todo para la definición de prioridades de inversiones pública y privada, las estimaciones de población desagregadas por grupos de edad. Esta información es aún más difícil de obtener que la magnitud total de la población, pues cuanto mayor desagregación mayor es la probabilidad de error en las estimaciones.

En el presente artículo se describen varios métodos de actualización de datos de población en los municipios del Estado de Rio Grande do Sul. Concretamente, se propone una metodología de estimación de población basada en el uso de variables sintomáticas.

La principal ventaja de las estimaciones de población —en comparación con los métodos tradicionales de proyección— es que permiten detectar los cambios de tendencias en el desplazamiento de la población ocurridos entre un censo y otro, por medio de información que puede considerarse indicativa de la tendencia del crecimiento

regional. Los métodos de proyección de tendencias sólo permiten observar cambios de tendencia del desplazamiento de la población de un área a otra cuando se analiza datos censales.

I. CONSIDERACIONES GENERALES

1. Reseña bibliográfica

El uso de variables sintomáticas para estimar el crecimiento de áreas menores aparece, muy incipientemente en Brasil, como indicador de cambios de tendencia del crecimiento de la población. Un estudio muy original en este campo fue el realizado por Irvin y Fialho, que estimaron la población del Distrito Federal en 1972 con el método compuesto, el método de componentes II y el método de tasas vitales (Irvin y Fialho, 1977). Los autores optan por el primero, que presenta los mejores resultados. Se utilizaron las siguientes variables sintomáticas: defunciones, nacimientos, matrimonios y matrícula escolar.

Según el inventario desarrollado por Laura Wong, solamente en Paraná y el Mato Grosso se aplicaron metodologías basadas en el uso de variables socioeconómicas como indicadores de crecimiento demográfico (Wong, 1987). En el primer caso se estimó la población urbana del Estado de Paraná, mientras que las estimaciones para el Estado de Mato Grosso fueron hechas con una mayor desagregación, a nivel de municipios. Aunque el objetivo de este estudio no es actualizar este inventario, se observó que se habían hecho otros cinco con estimaciones de población de áreas menores a partir de variables sintomáticas (Sawyer, 1969; Sawyer y Carvalho, 1981; Santa Catarina, 1989; IBGE, 1990 y Bega et al., 1991).

Donald Sawyer utiliza el método proporcional ponderado por el coeficiente de correlación entre la población y las variables sintomáticas para estimar la población de 16 municipios de Goiás en 1967 (Sawyer, 1969). Se pusieron a prueba trece indicadores agrícolas y luego se seleccionaron los que presentaban una correlación superior a 0.80 (arroz, maíz, ave, ganado y puerco) con el método proporcional ponderado. Debido a la alta correlación observada entre matrícula escolar y población, se hizo otra estimación utilizando solamente esa información. La estimación definitiva de población corresponde a la

media aritmética de las dos estimaciones (variables agrícolas y matrícula escolar).

Donald Sawyer y Magno de Carvalho utilizaron tres modelos de regresión múltiple para desagregar las proyecciones de población de las unidades de la Federación según condiciones de vivienda (Sawyer y Magno de Carvalho, 1981). En el primer modelo se establece una relación entre la población urbana de microrregiones homogéneas en 1970 con los datos pertinentes del censo industrial. Se seleccionaron las siguientes variables sintomáticas: media mensual de personas ocupadas, número de establecimientos, tamaño medio del establecimiento y participación del valor de la transformación industrial de la microrregión en el total del estado. En el segundo modelo se estima la población rural relacionándola con los datos del censo agropecuario referentes a microrregiones (área de cultivo temporal y permanente, proporción de la superficie total cultivada correspondiente a área de cultivo permanente, proporción de la superficie total cultivada correspondiente a área de pastoreo, total de personas empleadas, proporción de la producción total correspondiente a producción agrícola, indicador del nivel de actividad y relación entre nivel de actividad y total de personas ocupadas). En el tercer modelo se establece una relación entre la población urbana de las unidades de la Federación con las siguientes variables sintomáticas: número de camas de hospital, monto de la recaudación tributaria, número de electores, matrícula en la educación primaria y tiraje de periódicos. La estimación definitiva de la población urbana de las unidades federativas, se calculó sobre la base de la media aritmética de las diversas estimaciones del porcentaje de población urbana, una vez eliminados los valores discrepantes.

En el estudio realizado en Santa Catarina (Santa Catarina, 1989) se utilizó el aumento del número de electores de los municipios del estado en la década de 1980 para estimar la población intercensal.

El Instituto Brasileño de Geografía y Estadística (IBGE) utilizó la información sobre número de electores de las unidades de la Federación —recopilada en la encuesta nacional de hogares de 1988 para evaluar los datos del Tribunal Superior Electoral— y, posteriormente, para corregir distorsiones en la proyección de población para 1990 (IBGE, 1990).

En el Estado de Paraná también se utilizaron metodologías basadas en variables sintomáticas para corregir las proyecciones de población a nivel municipal de comienzos de la década (Bega et al., 1991). En esta

evaluación se emplearon las siguientes variables sintomáticas: matrícula escolar, número de electores, evolución del consumo de energía eléctrica y conexiones de agua potable. Para el análisis de la dinámica demográfica del área rural se recurrió a información sobre el sector agrícola (producción, estructura agraria, tecnificación y empleo).

Aunque en Brasil no se emplean sistemáticamente metodologías para estimaciones demográficas basadas en variables sintomáticas, algunos estudios publicados en varios países demuestran la eficacia y la viabilidad de esos métodos en la elaboración de estimaciones poblacionales para áreas de magnitud similar a un municipio.

En los Estados Unidos se comenzaron a realizar estimaciones demográficas con variables sintomáticas antes de los años cincuenta. Donald Bogue propuso el método de tasas vitales, que consiste en relacionar la natalidad y la mortalidad de una área menor con las del área mayor pertinente y hacer estimaciones de las tasas brutas de natalidad y mortalidad (Bogue, 1950). Debido a los efectos opuestos del envejecimiento y de la migración sobre las tasas de natalidad y mortalidad, Bogue proponía hacer estimaciones de población para áreas menores combinando las variables “nacimientos” y “defunciones”.

La experiencia que fueron adquiriendo los demógrafos estadounidenses les permitió perfeccionar las técnicas de estimación demográfica, utilizando varios métodos y variables sintomáticas para estimar con un grado de precisión aceptable, la población de áreas menores, incluso desagregada por grandes grupos de edad y sexo.

Crosetti y Schimitt (1956) ponen a prueba veinte series de estimaciones correspondientes a 39 condados del Estado de Washington en 1950, a partir de variables sintomáticas (nacimientos, defunciones, matrícula escolar, registro de automóviles, registro de electores y registro del seguro social). Aplicando el método de prorrateo y el método de razón censal (distribución proporcional), concluye que el uso de variables sintomáticas, aun con metodologías muy sencillas, ofrece estimaciones más precisas que los métodos matemáticos convencionales. La mejor estimación conseguida por el autor en esa prueba correspondió a la media de las estimaciones realizadas con el método de distribución proporcional, ponderada por el valor inverso de los coeficientes de variación ($1/CV$) de las respectivas variables sintomáticas.

Goldberg, Rao y Namboodiri (1964) comparan las estimaciones de la población de 83 condados del Estado de Michigan con los datos del censo de 1960. Utilizando los métodos de tasas vitales, el método

compuesto, el método de componentes II y la correlación de razón. Sobre la base de tres variables sintomáticas (nacimientos, defunciones y matrícula escolar), los autores concluyen que las estimaciones realizadas con el método de correlación de razón fueron las mejores (Goldberg, Rao y Namboodiri, 1964).

Peter Morrison presenta una evaluación comparativa de diferentes estudios y concluye que la eficacia del método varía de un área a otra, dependiendo de la desagregación geográfica, el grado de urbanización y la movilidad, entre otras variables, por lo que aboga por aplicarlos con flexibilidad y adaptándolos a los datos disponibles (Morrison, 1971).

2. Objetivos y descripción del estudio

Al parecer, el empleo de información socioeconómica para estimar el crecimiento de la población de áreas menores tenderá a extenderse en Brasil, en parte porque la recopilación de información regional, sistematizada por instituciones estatales, ha mejorado notablemente en los últimos años. Además, el actual desinterés del gobierno federal por la recopilación y difusión de ciertos datos estadísticos obligará a los investigadores a aplicar métodos alternativos.

Rio Grande do Sul cuenta con condiciones que hacen posible la aplicación de metodologías de estimación de población, dado que ya existen registros de información sistematizados, a nivel de municipio, que pueden utilizarse como indicadores del crecimiento poblacional. La principal ventaja de usar información estructurada de esta manera estriba en el hecho de que, por responder a los criterios empleados en el desarrollo de bancos de datos, su actualización y recolección es ágil, dinámica y confiable, características imprescindibles para el uso de variables sintomáticas con fines de estimación de población.

Además, y al contrario de lo que ocurre con las metodologías empleadas para proyecciones demográficas,² el dinamismo del proceso de estimación de la población hace que se vaya perfeccionando cada vez que se añade nueva información relevante al acervo de datos disponibles. En este sentido, la mejora de las estimaciones está íntimamente ligada a la calidad de la información utilizada.

2 En este caso, el término "estimación" se refiere concretamente a la aplicación del método de inferencia estadística, en tanto que "proyección" se refiere a la extrapolación de la tendencia poblacional observada en décadas anteriores.

El objetivo de este proceso es utilizar información socioeconómica para estimar el volumen de la población de los municipios del Estado de Rio Grande do Sul a lo largo de toda la década de 1990. El presente estudio consiste en una evaluación de la eficacia de diferentes variables sintomáticas y distintos métodos de distribución de la población del estado entre sus municipios, sobre la base de los resultados del censo demográfico de 1991.

3. Metodología

Para evaluar los resultados obtenidos con el empleo de distintas metodologías de estimación de la población se puede recurrir al cálculo de estimaciones correspondientes a los años en que se han realizado censos. En síntesis, lo que se propone es un análisis de la precisión de las estimaciones realizadas, de acuerdo a distintas modalidades de distribución de la población del estado entre sus municipios.

En el presente estudio se evalúan los métodos utilizados en estimaciones de población por medio de variables representativas. El universo de estimación, establecido previamente, es la población del Estado. El cálculo de estimaciones demográficas con este método no sólo tiene la ventaja de que la suma de la población de todos los municipios corresponde a la población total del Estado sino que, además, ofrece un grado de precisión que no se ve afectado por la cobertura de las variables utilizadas, que son muy similares (prácticamente iguales) en todas las áreas.

Fuentes de información

Para realizar estimaciones demográficas se necesita contar con los datos sobre población recopilados en el último censo demográfico disponible y un análisis de las variables sintomáticas. Se seleccionaron aquellas variables representativas del volumen de población que contaban con datos correspondientes a todos los años considerados y todos los municipios. La primera restricción que delimita la unidad de medida de la variable sintomática, es decir la unidad de medida de todos los datos, es el número de personas. La segunda elimina las variables que, a pesar de ser ampliamente utilizadas en la bibliografía consultada, no están sistematizadas ni han sido recopiladas con periodicidad anual en todos los municipios del Estado. En este caso se

tomó en cuenta el número de consumidores de energía eléctrica (la empresa estadual de electricidad no cubre todos los municipios del Estado) y el número de personas empleadas en la industria y comercio (en 1991, el Ministerio de Hacienda proporcionó información al respecto sobre todos los municipios del Estado, pero en 1980 sólo entregó datos sobre los 50 municipios de mayor magnitud).

En vista de dichas limitaciones, se seleccionaron las siguientes variables sintomáticas: número de electores registrados, matrícula inicial en la educación básica (primer a octavo grado), número de nacidos vivos registrados por lugar de residencia de la madre, número de defunciones registradas por lugar de residencia. Las estadísticas vitales corresponden al total registrado en el año. La matrícula escolar comprende al total de alumnos matriculados 30 días después de iniciado el año escolar. El número de electores corresponde a brasileños mayores de 16 años de edad, clasificados en votantes que ejercen su derecho a voto y votantes que se abstienen.

Las fuentes de estos datos son el Instituto Brasileño de Geografía y Estadística (IBGE), la Secretaría de Educación del Estado de Rio Grande do Sul y el Tribunal Regional Electoral.

Con el objeto de eliminar posibles fluctuaciones aleatorias en las variables, se calculó la media de los datos correspondientes a los años 1988, 1989 y 1990. Esta discrepancia entre el período cubierto por las variables sintomáticas y la población se debe al retraso sistemático en la divulgación de los datos, sobre todo de las estadísticas vitales. Para el año 1980 también se calculó la media de la información sintomática de los años 1979, 1980 y 1981.

En los cuadros 1 y 2 se presentan los coeficientes de correlación de esas variables con los datos censales sobre población de las 223 agrupaciones geográficas incluidas en el estudio.

Cuadro 1
**COEFICIENTE DE CORRELACIÓN DE PEARSON ENTRE
 LA POBLACIÓN DE 1991 Y LAS VARIABLES SINTOMÁTICAS
 (MEDIA DE 1988, 1989 Y 1990)**

Grupos de edad	Nacidos vivos	Defunciones	Número de electores	Matrícula en educación básica
De 0 a 6 años	0.995	0.981	0.987	0.996
De 7 a 14 años	0.997	0.987	0.992	0.998
De 15 a 49 años	0.999	0.991	0.997	0.997
De 50 y más años	0.990	0.998	0.998	0.990
Total	0.999	0.993	0.997	0.998

Fuente: Los datos brutos provienen del Instituto Brasileño de Geografía y Estadística, la Secretaría de Educación del Estado de Rio Grande do Sul y el Tribunal Regional Electoral.

Cuadro 2
**COEFICIENTE DE CORRELACIÓN DE PEARSON ENTRE
 LA POBLACIÓN DE 1980 Y LAS VARIABLES SINTOMÁTICAS
 (MEDIA DE 1979, 1980 Y 1981)**

Grupos de edad	Nacidos vivos	Defunciones	Número de electores	Matrícula en educación básica
De 0 a 6 años	0.998	0.987	0.992	0.988
De 7 a 14 años	0.996	0.984	0.988	0.991
De 15 a 49 años	0.995	0.991	0.998	0.984
De 50 y más años	0.987	0.995	0.997	0.978
Total	0.997	0.993	0.998	0.987

Fuente: Los datos brutos provienen del Instituto Brasileño de Geografía y Estadística, la Secretaría de Educación del Estado de Rio Grande do Sul y el Tribunal Regional Electoral.

Métodos de estimación de la población de áreas menores

En cuanto al trato dado a las variables, se aplicaron varios métodos para hacer las estimaciones: el de distribución por prorrateo, el de distribución proporcional (Crosetti y Schimitt, 1956), el de tasas vitales (Bogue, 1950), una variación del método compuesto de Bogue y Duncan (descrito en Morrison, 1971) y el de correlación de razón (Goldberg, Rao y Namboodiri, 1964). Los procesos pertinentes se describen a continuación.

$$P_{h,t} = \frac{S_{h,t}}{S_{T,t}} \times P_{T,t}$$

a) Método de distribución por prorrateo

En la distribución por prorrateo se supone que la razón entre la población de cada municipio y la del Estado es igual a la razón correspondiente a la variable sintomática:

donde:

$P_{h,t}$	es la estimación de la población del municipio h en el año t ;
$S_{h,t}$	es el valor de la variable sintomática S para el municipio h en el año t ;
$P_{T,t}$	es la estimación de la población para el total del Estado en el año t ;
$S_{T,t}$	es el valor de la variable sintomática S para el total del Estado en el año t ;

En esta ecuación, $P_{T,t}$ corresponde a la población según el censo de 1991, y $S_{h,t}$ y $S_{T,t}$ corresponden a la media de las variables sintomáticas (nacidos vivos, defunciones, electores y matrícula escolar en la educación primaria) de los años 1988, 1989 y 1990.

b) Método de distribución proporcional

La aplicación de este método se basa en el supuesto de que la población crece en la misma proporción que la variable sintomática.

$$P_{h,t} = \frac{P_{h,0} \times \frac{S_{h,t}}{S_{h,0}}}{\sum \left(P_{h,0} \times \frac{S_{h,t}}{S_{h,0}} \right)} \times P_{T,t}$$

El factor $\{P_{T,t} / \sum [P_{h,0} \times (S_{h,t} / S_{h,0})]\}$ permite adaptar la fórmula para que haya consistencia con la población total del estado en el año t e impide que las estimaciones se vean influenciadas por cambios en la cobertura de las variables sintomáticas, siempre que éstas presenten la misma tendencia en todos los municipios.

Los datos referentes al año t son los definidos en la ecuación correspondiente al método de distribución por prorratio; los del año 0 equivalen a la población del censo de 1980 y la media de las variables sintomáticas centradas en 1980³.

Dado que en la aplicación de ese método se utilizan dos puntos dados en el tiempo, fue necesario compatibilizar las áreas geográficas de 1980 y 1991. Los datos considerados en este estudio corresponden a áreas mínimas de comparación (223 observaciones), que pueden ser municipios o agrupaciones de municipios.

c) Método de tasas vitales

El método de tasas vitales propuesto por Bogue consiste en estimar la población de los municipios utilizando solamente información proveniente de las estadísticas vitales (Bogue, 1950). Para aplicar este método se necesita conocer el número de nacimientos y defunciones en el año base y en el año considerado para la estimación, en cada municipio, así como la población del área mayor y las tasas brutas de natalidad y de mortalidad.

Donald Bogue parte del supuesto de que hay una relación inversa entre las variaciones anuales de las tasas de natalidad y de mortalidad de los municipios y de que la desviación proporcional de esas tasas con respecto a las tasas estatales se mantiene constante a lo largo del tiempo. El autor sostiene que es posible hacer una estimación fidedigna de la población de áreas menores mediante la combinación de estimaciones realizadas usando la razón entre las tasas municipales y las del área mayor. El modelo matemático utilizado en este caso puede describirse de la siguiente manera:

3 Los datos sobre nacimientos, defunciones y matrícula corresponden a los años 1979, 1980 y 1981. En el caso de los electores se consideró la media de 1979 y 1982 (años de elecciones).

$$P_{h,t} = a \times \left(\frac{B_{h,t}}{b_{h,t}} \right) + (a-1) \times \left(\frac{D_{h,t}}{d_{h,t}} \right)$$

donde:

$$b_{h,t} = \frac{\frac{B_{h,0}}{P_{h,0}}}{\frac{B_{T,0}}{P_{T,0}}} \times \left(\frac{B_{T,t}}{P_{T,t}} \right) \qquad d_{h,t} = \frac{\frac{D_{h,0}}{P_{h,0}}}{\frac{D_{T,0}}{P_{T,0}}} \times \left(\frac{D_{T,t}}{P_{T,t}} \right)$$

- a es el coeficiente de ponderación entre las dos estimaciones (en este caso se adoptó $a=0,50$ y $a=0,54$)
- $B_{h,t}$ y $B_{h,0}$ son a los nacidos vivos del municipio h , en los años t y 0 , respectivamente,
- $B_{T,t}$ y $B_{T,0}$ son el total de nacidos vivos en el Estado en los años t y 0 , respectivamente,
- $D_{h,t}$ y $D_{h,0}$ son las defunciones en el municipio h , en los años t y 0 , respectivamente,
- $D_{T,t}$ y $D_{T,0}$ son el total de defunciones en el Estado en los años t y 0 , respectivamente,
- $b_{T,t}$ es la tasa de natalidad total del Estado en el año t , y
- $d_{T,t}$ es la tasa de mortalidad total del Estado en el año t .

En este caso se hicieron dos estimaciones. En la primera se utilizó el mismo coeficiente de ponderación para las estimaciones realizadas con la variable “nacimientos” y “defunciones” (0.50), mientras en la segunda se usó el coeficiente de ponderación propuesto por Bogue, por el inverso del coeficiente de variación de las respectivas variables sintomáticas ($1/CV_b=0.4292$ para nacimientos y $1/CV_d=0.3837$ para defunciones).

d) Variación del método compuesto

El método compuesto propuesto por Bogue y Duncan consiste en hacer estimaciones de población desagregadas por grupos de edad y por sexo (Morrison, 1971). Para estimar la población de 45 años y más, se utilizan los datos sobre defunciones; para la población femenina de 15 a 44 años y la población de menos de 5 años, los datos sobre nacimientos

por edad de la madre, y para la población de 5 a 14 años, los datos sobre matrícula escolar. Dado que el desglose de las estadísticas vitales por grupos de edad, sobre todo de las defunciones, es poco confiable en áreas pequeñas, se consideró el valor total de las variables sintomáticas, sin desglosarlas. La variación que se propone consiste en estimar por separado la población de 0 a 6 años, a partir de la cifra sobre nacidos vivos; la de 7 a 14 años, sobre la base de la matrícula escolar inicial en la educación primaria; la de 15 a 49 años, utilizando el número de electores, y la de 50 años y más, a partir del número de defunciones. Las ecuaciones utilizadas en el modelo propuesto son:

$$P_{h,t}(0-6) = \frac{P_{h,0}(0-6) \times \frac{B_{h,t}}{B_{h,0}}}{\sum \left[P_{h,0}(0-6) \times \frac{B_{h,t}}{B_{h,0}} \right]} \times PT_{,t}(0-6)$$

$$P_{h,t}(7-14) = \frac{P_{h,0}(7-14) \times \frac{M_{h,t}}{M_{h,0}}}{\sum \left[P_{h,0}(7-14) \times \frac{M_{h,t}}{M_{h,0}} \right]} \times PT_{,t}(7-14)$$

$$P_{h,t}(15-49) = \frac{P_{h,0}(15-49) \times \frac{E_{h,t}}{E_{h,0}}}{\sum \left[P_{h,0}(15-49) \times \frac{E_{h,t}}{E_{h,0}} \right]} \times PT_{,t}(15-49)$$

$$P_{h,t}(50+) = \frac{P_{h,0}(50+) \times \frac{D_{h,t}}{D_{h,0}}}{\sum \left[P_{h,0}(50+) \times \frac{D_{h,t}}{D_{h,0}} \right]} \times PT_{,t}(50+)$$

donde:

$P_{h,t}(x-y)$ es la población de x ($x= 0, 7, 15$) a y ($y= 6, 14, 49$) años de edad del municipio h , en el año t ;

$P_{h,0}(x-y)$ es la población de x a y años de edad del municipio h , en el año 0;

$B_{h,t}$ y $B_{h,0}$	son los valores de la variable sintomática “nacimientos”, definidos mediante la aplicación de los métodos anteriores;
$P_{T,t}(x-y)$	es la población de x a y años de edad de todo el estado en el año t ;
$M_{h,t}$ y $M_{h,0}$	son los valores de la variable sintomática “matrícula escolar” del municipio h en los años t y 0 , respectivamente;
$P_{h,t}(50+)$	es la población de 50 años y más del municipio h en el año t ;
$E_{h,t}$ y $E_{h,0}$	son los valores de la variable sintomática “electores” del municipio h en los años t y 0 , respectivamente, y
$D_{h,t}$ y $D_{h,0}$	son los valores de la variable sintomática “defunciones”, definidos mediante la aplicación de los métodos anteriores.

e) Método de correlación de razón

El método de correlación de razón se basa en el supuesto de que la población crece en la misma proporción que las variables sintomáticas seleccionadas como representativas de variación del tamaño de la población. Sobre la base de este supuesto, se formula un modelo de regresión en que se utiliza el método de regresión múltiple para estimar los parámetros; con este fin, se establece una relación entre las razones intercensales de la proporción de población de los municipios y del Estado (variable dependiente) y las razones existentes entre las proporciones de las variables sintomáticas de los municipios y del Estado (variables independientes). En este estudio se aplicó el modelo al crecimiento intercensal entre 1980 y 1991.

Dado que este modelo se utilizará también para estimar la población desglosada por grupos de edad, en la sección siguiente se lo describe más en detalle.

f) Método de correlación de razón para estimar la población por grupos de edad

Tomando en cuenta, por una parte, la necesidad de contar con estimaciones de población desglosadas por grupos de edad y, por otra,

la dificultad de lograr un grado de precisión aceptable, se optó por usar la estructura de la población por grupos de edad dividida solamente en cuatro grupos.

Tomando en cuenta la demanda de planificadores y demógrafos, se seleccionaron grupos de edad más amplios: población infantil (de 0 a 6 años), niños en edad de cursar estudios básicos (de 7 a 14 años), población en edad reproductiva y productiva (15 a 49 años) y tercera edad. La población en edad activa debería, en principio, abarcar hasta los 60 o 65 años, pero se optó por hacer el corte a los 50 años, considerando que el grupo de 50 a 65 años sería muy reducido en los municipios, lo que limitaría la significación de las estimaciones, sobre todo en los municipios de menor tamaño.

Ejemplo de modelo estimado:

$$Y_{ih} = a_{i0} + a_{i1} \times X_{1h} + a_{i2} \times X_{2h} + a_{i3} \times X_{3h} + a_{i4} \times X_{4h}$$

donde:

$$Y_{ih} = \frac{\left(\frac{P_{ih,t}}{P_{iT,t}} \right)}{\left(\frac{P_{ih,0}}{P_{iT,0}} \right)} \quad X_{jh} = \frac{\left(\frac{S_{jh,t}}{S_{jT,t}} \right)}{\left(\frac{S_{jh,0}}{S_{jT,0}} \right)}$$

En la estimación del modelo se utilizó la siguiente información:

$P_{ih,t}$ = población del grupo de edad i , del municipio h , en el año censal 1991;

$P_{iT,t}$ = población del grupo de edad i , del Estado de Rio Grande do Sul, en el año censal 1991;

$P_{ih,0}$ = población del grupo de edad i , del municipio h , en el año censal 1980;

$P_{iT,0}$ = población del grupo de edad i , del Estado de Rio Grande do Sul, en el año censal 1980;

$S_{1h,t}$ = promedio de electores del municipio h en los años 1988, 1989 y 1990;

- $S_{1T,t}$ = promedio de electores del Estado en los años 1988, 1989 y 1990;
- $S_{1h,0}$ = promedio de electores del municipio h en los años 1979 y 1981;
- $S_{1T,0}$ = promedio de electores del Estado en los años 1979 y 1981;
- $S_{2h,t}$ = promedio de jóvenes matriculados en educación básica en el municipio h en los años 1988, 1989 y 1990;
- $S_{2T,t}$ = promedio de jóvenes matriculados en educación básica en el Estado en los años 1988, 1989 y 1990;
- $S_{2h,0}$ = promedio de jóvenes matriculados en educación básica en el municipio h en los años 1979, 1980 y 1981;
- $S_{2T,0}$ = promedio de jóvenes matriculados en educación básica en el Estado en los años 1979, 1980 y 1981;
- $S_{3h,t}$ = promedio de nacidos vivos en el municipio h en los años 1988, 1989 y 1990;
- $S_{3T,t}$ = promedio de nacidos vivos en el Estado en los años 1988, 1989 y 1990;
- $S_{3h,0}$ = promedio de nacidos vivos en el municipio h en los años 1979, 1980 y 1981;
- $S_{3T,0}$ = promedio de electores del Estado en los años 1979, 1980 y 1981;
- $S_{4h,t}$ = promedio de defunciones en el municipio h en los años 1988, 1989 y 1990;
- $S_{4T,t}$ = promedio de defunciones en el Estado en los años 1988, 1989 y 1990;
- $S_{4h,0}$ = promedio de defunciones en el municipio h en los años 1979, 1980 y 1981;
- $S_{4T,0}$ = promedio de defunciones en el Estado en los años 1979, 1980 y 1981;

En cuanto a la división territorial utilizada en el modelo de regresión, cabe señalar que, cuando se utilizaron dos años censales en la

estimación (1980 y 1991) hubo que compatibilizar las áreas geográficas de 1980 y 1991. Al igual que en el caso del método de distribución proporcional, los datos corresponden a las áreas de menor tamaño. En el cuadro 3 se indican los parámetros estimados.

Cuadro 3
COEFICIENTES DE REGRESIÓN DEL MODELO
DE CORRELACIÓN DE RAZÓN

Grupos de edad	Constante	Nacimientos	Defunciones	Electores	Matrícula
Menores de 6 años	0.0101	0.4881	0.0011	-0.0606	0.5653
De 7 a 14 años	-0.1048	0.0945	0.0022	0.0648	0.9025
De 15 a 49 años	0.0260	0.2132	-0.0799	0.4098	0.4536
De 50 y más años	0.3981	0.0367	0.1687	0.2451	0.1557
Total	0.1062	0.2312	0.0256	0.0968	0.5289

Fuente: Elaboración propia.

La principal ventaja del método de correlación de razón es que no es necesario que la información correspondiente a un área determinada haya sido considerada en la estimación de los parámetros del modelo de regresión. A pesar de que en la elaboración del modelo se trabajó con agrupaciones de municipios, es posible hacer estimaciones sobre todos los municipios del Estado y sobre los años posteriores, dado que las variables sintomáticas utilizadas en el modelo responden a esa división territorial. Por este motivo, el método de correlación de razón es imprescindible para la estimación de la población de municipios creados con posterioridad al último censo.

Suponiendo que la relación estimada entre las razones de proporción de la población y de las variables sintomáticas siga siendo válidas para años posteriores al último censo, se estima la población de los municipios desglosada por grupos de edad, mediante las siguientes ecuaciones:

$$Y_{ih} = \frac{\frac{P_{ih,t+n}}{P_{iT,t+n}}}{\frac{P_{ih,t}}{P_{iT,t}}} \qquad X_{jh} = \frac{\frac{S_{jh,t+n}}{S_{jT,t+n}}}{\frac{S_{jh,t}}{S_{jT,t}}}$$

donde:

- $P_{ih,t+n}$ es la población estimada del grupo de edad i , del municipio h , n años después del último censo;
- $P_{iT,t+n}$ es la población proyectada del grupo de edad i , de todo el Estado, n años después del último censo;
- $P_{ih,t}$ es la población del grupo de edad i , del municipio h , en la fecha del último censo;
- $P_{iT,t}$ es la población del grupo de edad i , de todo el Estado, en la fecha del último censo;
- $S_{jh,t+n}$ es el valor de la variable sintomática j , del municipio h , n años después del último censo;
- $S_{jT,t+n}$ es el valor de la variable sintomática j , de todo el Estado, n años después del último censo;
- $S_{jh,t}$ es el valor de la variable sintomática j , del municipio h , en el año del último censo;
- $S_{jT,t}$ es el valor de la variable sintomática j , para el total del Estado, en el año del último censo.

4. Evaluación de los resultados

El análisis de los resultados se basa en una comparación de las estimaciones para el año de 1991 con los datos pertinentes del censo demográfico de 1991. Para el análisis de las estimaciones se utiliza el error relativo, que puede describirse mediante la siguiente ecuación:

$$e_h = \left[\frac{(PC_h - PE_h)}{PC_h} \right] \times 100$$

donde:

- e_h es el error relativo de la estimación de la población del municipio h ;
- PC_h es la población censada del municipio h , y
- PE_h es la población estimada del municipio h .

En la evaluación de las estimaciones se toman en cuenta los siguientes valores referentes a errores relativos: media, y desviación estándar y porcentual entre -10% y 10%. Los dos últimos indican la eficacia del método empleado (cuanto menor la variabilidad de los errores, más eficaz es la metodología), mientras que la primera revela la eficacia de las estimaciones (las que presentan errores medios próximos a cero son las mejores). En el cuadro 4 se presenta un resumen de las estadísticas empleadas para estimar la población total.

En el caso del método de distribución por prorratio, la variable que posibilitó las mejores estimaciones fue la matrícula escolar, con un error medio de -1.57%; en cambio, las peores estimaciones son las basadas en la variable “defunciones” (error medio de -12.38). Mientras la matrícula escolar ofrece las estimaciones más precisas (menor error medio), el número de electores permite el más alto grado de eficacia (desviación estándar de 9.94 y cerca del 70% de los municipios con error relativo de -10% a 10%).

Las estimaciones realizadas sobre la base del método de tasas vitales presentan errores medios menores que las estimaciones realizadas conforme al método de distribución por prorratio, cuando se consideran nacimientos y defunciones, independientemente de la ponderación utilizada. La combinación de esas variables permite una precisión y una eficacia mucho mayores que la relación univariada de esa información.

En las estimaciones basadas en el método de distribución proporcional las variables sintomáticas referentes a las estadísticas vitales presentan el menor índice de error medio relativo: 0.16% en el caso de los nacimientos y 4.49% en el de las defunciones. En cambio, cuando se emplea el método de distribución proporcional las variables sintomáticas “electores” y “matrícula escolar” presentan sesgos más altos que cuando se aplica el método de distribución por prorratio. En cuanto a la eficacia de las estimaciones, cabe destacar que, con excepción de la variable “electores”, todas las demás permitan hacer estimaciones más eficaces cuando se emplea el primero de estos métodos en lugar del segundo. En comparación con el método de tasas vitales, la precisión del método de distribución proporcional sólo es superior en el caso de las estimaciones basadas en la variable “matrícula escolar”.

Cuadro 4
RESUMEN DE LAS ESTADÍSTICAS SOBRE ERRORES
RELATIVOS. 1991

Método	Error medio	Desviación estándar	Error máximo	Error mínimo	Coefficiente de variación	Porcentaje de municipios con error de -10% a 10%
Correlación de razón	-0.31	5.50	22.44	-13.67	0.06	91.93
Método compuesto	2.24	8.24	27.84	-21.69	0.27	77.58
Distribución proporcional						
Nacimientos	0.16	15.85	48.84	-47.03	0.01	53.36
Defunciones	4.49	18.80	75.86	-38.35	0.24	49.78
Electores	5.47	13.43	52.06	-28.69	0.41	59.64
Matrícula	-5.52	9.01	17.97	-31.88	0.61	68.16
Tasas vitales						
Igual ponderación	2.31	12.28	53.82	-28.20	0.19	68.16
Ponderación 1/CV	2.19	12.18	52.59	-28.52	0.18	68.16
Distribución por prorrateo						
Nacimientos	-4.41	23.37	92.96	-92.38	0.19	46.85
Defunciones	-12.38	24.27	59.30	-94.33	0.51	33.03
Electores	4.25	9.94	54.82	-23.63	0.43	69.67
Matrícula	-1.57	14.66	41.09	-39.89	0.11	47.15

Fuente: Elaboración propia.

Cuando los resultados obtenidos mediante el método compuesto se ajustan al total del estado, muestra una eficacia superior a todos los demás métodos analizados (77.58% de errores relativos entre -10% y 10%). Cabe destacar que, a pesar de que el método compuesto presenta un grado razonable de eficacia en las estimaciones del total de la población, si ésta se desglosa por grupos de edad los resultados no son satisfactorios. En el caso de todos los grupos de edad, el método de correlación de razón ofrece resultados mucho mejores que los obtenidos mediante la aplicación del método compuesto.

El método de correlación de razón fue el que produjo los mejores resultados, dado que presentan errores superiores al 10% o inferiores a -10% en menos de un 10% de los municipios. Además, en cuanto al sesgo de las estimaciones, el error medio relativo es muy cercano a cero (-0.28%). Las estimaciones realizadas con el método de distribución proporcional, utilizando la variable sintomática “nacimientos”, presentan un error medio inferior al método de correlación de razón. En cambio, la proporción de estimaciones con error de menos del 10% es mucho más alta cuando se emplea el método de correlación de razón que el de distribución proporcional considerando la variable “nacimientos”.

En términos generales, las estimaciones realizadas con el método de correlación de razón, desglosadas por grupos de edad, son las que presentan los resultados más satisfactorios, con errores medio inferiores al 3%. En las estimaciones no ajustadas al total del Estado, todos los grupos de edad presentan un error medio positivo; en cambio, cuando se ajustan al total solamente la población de 7 a 14 años presenta un error medio positivo.

En el caso de las estimaciones no ajustadas al total del Estado, el grupo de 15 a 49 años es el que presenta resultados menos precisos (menor error medio). El ajuste de las estimaciones al total del Estado mejora la precisión y la eficacia de las estimaciones de la población 0 a 6 años y de 50 y más años, y la reduce en los demás grupos de edad.

En el análisis del patrón de errores de las estimaciones de población por tamaño de municipios, se observa que, en general, éstos son menores en los municipios más extensos. El margen de error se reduce notablemente cuando se emplea el método de distribución por prorratio aplicado a los nacimientos; en tal caso, el 87.50% de los municipios de mayor tamaño muestra un error de -10% a 10%, lo que

se compara con un 33.13% y un 57.14% de los municipios menores e intermedios, respectivamente; cuando se emplea el método de tasas vitales, el 90% de los municipios de mayor tamaño presenta un error medio de -10% a 10%, en comparación con alrededor de un 60% a 70% en los municipios menores e intermedios.

En relación con el error medio relativo de las estimaciones de población por grupos de edad, con excepción del grupo de 7 a 14 años, las estimaciones no ajustadas al total del Estado también son menos precisas en los municipios de menor tamaño. El ajuste al total del estado invierte esta situación, por lo que cabe destacar que, a pesar de el ajuste al total del Estado aumenta el error medio de las estimaciones de los municipios mayores, las diferencias no son muy significativas. En cambio, el ajuste de las estimaciones pertinentes permite un grado mucho mayor de precisión.

5. Conclusiones

A pesar de las limitaciones observadas en relación con la disponibilidad de datos sobre las variables sintomáticas, la creación de nuevos municipios y la periodicidad de la información⁴, la aplicación de los modelos de estimación de población permite llegar a las siguientes conclusiones:

- La eficacia de las estimaciones basadas en el método de distribución por prorrateo, dada por la proporción de municipios con error de -10% a 10%, está directamente vinculada a la correlación entre la variable sintomática y la población. En el caso de los municipios con más de 100 000 habitantes, en promedio el uso de la variable sintomática “número de electores” suele conducir a una significativa subestimación de la población (error relativo medio de -6.89%), mientras las demás conducen a una sobrestimación. En el caso de los municipios pequeños, con menos de 10 000 habitantes, el número de electores suele conducir a una sobrestimación (error relativo medio de 7.15%), mientras las

4 Luego de aplicar varios métodos de estimación para el área metropolitana de Hawai, Schmitt y Grier (1966) llegaron a la conclusión de que variables útiles en la década de 1940 dejaron de serlo en el decenio siguiente.

demás variables tienden a traducirse en una subestimación. Esto indica que la estratificación simple probablemente permita hacer estimaciones mucho más precisas, basadas en el método de distribución por prorrateo⁵.

- A diferencia de lo que ocurre con el método de distribución por prorrateo, el de distribución proporcional permite estimaciones más precisas a partir de los datos sobre nacimientos y defunciones, y menos precisas cuando se emplean las variables “electores” y “matrícula escolar”. El mayor grado de precisión que se logra en el primer caso es superior a la pérdida de precisión en el segundo. En este sentido, considerando el conjunto de variables utilizadas, el método de distribución proporcional es bastante más preciso que el de distribución por prorrateo.
- El método de correlación de razón ofrece resultados muy eficaces: más del 90% de los municipios presentan un error relativo de -10% a 10%, lo que significa que solamente en un caso se observa un error fuera del intervalo -20% y 20%.
- Los errores significativos en el caso de poblaciones de menor tamaño son inherentes al proceso de estimación de “universos” (en el sentido que se utiliza el término en teoría de muestreo) menores. Además, el empleo de una medida de error relativa en lugar de absoluta es una forma de dar prioridad a la precisión de las estimaciones correspondientes a áreas mayores.
- En todo proceso de estimación la precisión depende del mantenimiento a lo largo del tiempo de la tendencia de crecimiento registrada en el pasado reciente. Según Mandell y Taymann, a pesar de que la inestabilidad temporal de los coeficientes influye en el grado de error de las estimaciones, la influencia no es tan marcada como se afirma en la literatura sobre el tema (Mandell y Taymann, 1982).

A pesar de que la contribución metodológica de este estudio es limitada, se considera que los resultados obtenidos demuestran que es posible realizar estimaciones de población bastante precisas sobre la

5 Esto significa que, en lugar de distribuir la población total del Estado entre todos los municipios, éstos se agrupan de acuerdo a un determinado criterio (tamaño, grado de urbanización y tasas de crecimiento, entre otras) y se aplica la metodología independientemente a estos grupos.

base de datos dados a conocer anualmente, sin necesidad de recurrir a metodologías muy sofisticadas.

El método de correlación de razón parece ser el más indicado para estimar la población de los municipios del estado de Rio Grande do Sul. Además de su eficacia y de permitir estimaciones más precisas, tiene la ventaja de permitir que se lo aplique para estimar la población de municipios creados después de la fecha del último censo.

En conclusión, cabe señalar que, dadas las limitaciones para la ejecución de este estudio, se considera que los resultados son satisfactorios y que hay muchas posibilidades de perfeccionarlo en términos de precisión. En este sentido, el análisis de los errores desglosados por tamaño de la población indica que probablemente los resultados podrían ser más eficaces si los municipios se estratificaran por tamaño. Se determinó que la estratificación en función del tamaño en lugar de las tasas de crecimiento, que es la variable utilizada por Martin y Serow (1978), es más adecuada, por el hecho de que hay menos probabilidades de variación de una década a otra si se considera el tamaño en vez del ritmo de crecimiento.

BIBLIOGRAFÍA

- Bega, M.T., M.V.D. Soares y R.N. Azzolini (1991), "População paranaense nos anos 80: evolução e distribuição espacial", *Análise de conjuntura*, vol. 13, Nº 3, Curitiba, Instituto Paranaense de Desenvolvimento Económico y Social (IPARDES).
- Bercovich, Alicia y M.G.G. Morell (1992), "Notas sobre as informações da sinopse preliminar do censo demográfico de 1991", *Censo Demográfico de 1991: análises preliminares*, vol. 2, Rio de Janeiro, Directoria de Pesquisa, Instituto Brasileiro de Geografia y Estadística (IBGE), octubre.
- Bogue, Donald. (1950), "A technique for making extensive population estimates", *Journal of the American Statistical Association*, vol. 45, Nº 250.
- Carvalho, José Alberto Magno y Sílvia de Menezes Gama Pinheiro (1986), *Fecundidade e mortalidade no Brasil 1970/1980*, Belo Horizonte, Centro de Desenvolvimento y Planificación Regional (CEDEPLAR), Universidad Federal de Minas Gerais (UFMG).
- Crosetti, Albert y Robert Schmitt (1956), "A method of estimating the intercensal population of counties", *Journal of American Statistical Association*, Nº 51.
- Goldberg, David, V.R. Rao y N.K. Namboodiri (1964), "A test of the accuracy of ratio correlation population estimates", *Land Economics*, vol. 40, Nº 1.
- IBGE (Instituto Brasileiro de Geografia e Estatística) (1992), *Anuário estatístico do Brasil - 1992*, vol. 52, Rio de Janeiro.
- (1990), *Estimativa da população residente, em 01-07-90. Segundo os municípios e a situação político administrativa vigente em 15-11-89, com indicação dos coeficientes T.C.U relativos ao F.P.M*, Rio de Janeiro, inédito.
- Irvin, Richard y Eduardo Fialho (1977), "Estimativa provisória da população do Distrito Federal: 15 de novembro de 1972", *Revista Brasileira de Estatística*, Nº 35, Nº 146.
- Martin, Julia H. y Willian J. Serow (1978), "Estimating demographic characteristics using the ratio-correlation method", *Demography*, vol.13, Nº 2, febrero.
- Mandell, Marylou y Jeffrey Tayman (1982), *Measuring temporal stability in regression models of population estimation*, vol. 19, Nº 1, febrero.

- Morrison, Peter A. (1971), "Demographic information for cities: a manual for estimating and projecting local population characteristics", Rand Report, Nº R-618-HUD, Santa Monica, Rand Corp.
- Santa Catarina (1989), *Estudo da evolução populacional segundo os municípios 1970-2010*, Florianópolis, Secretaria de Estado de Coordenação e Planejamento, Subsecretaria de Estudos Geográficos e Estatísticos.
- Sawyer, Donald R. (1969), "Methods of estimating current population size and distribution", *Penetration Roads and Settlement on the Belem-Brasilia Highway*, Harvard, Department of Social Relations, Harvard College.
- Sawyer, Diana Oya y José Alberto Magno de Carvalho (1981), *Projeções da população brasileira: 1980-2000*, Belo Horizonte, Centro de Desarrollo y Planificación Regional (CEDEPLAR)/ Universidad Federal de Minas Gerais (UFMG)/ Ministerio de Transporte.
- Schmitt, Robert y J. Grier (1966), "A method of estimating the population of minor civil divisions", *Rural Sociology*, vol.31, Nº 3.
- Taschner, Suzana P. y Lucia M. Bógus (1986), "Mobilidade espacial da população brasileira: aspectos e tendências recentes", *Revista Brasileira de Estudos Populacionais*, vol. 3, Nº 2, Campinas, julio-diciembre.
- Teixeira Jardim, Maria de Lourdes (1995), *Utilização de variáveis sintomáticas para estimar a distribuição espacial de populações. Aplicação aos Municípios do Rio Grande do Sul*, Belo Horizonte, Centro de Desarrollo y Planificación Regional (CEDEPLAR)/ Federación de Órganos para Asistencia Social y Educacional (FASE)/ Universidad Federal de Minas Gerais (UFMG).
- Wong, Laura R. (1987), "Situação das projeções de população a nível de unidades federativas e algumas grandes regiões", *Futuro da população brasileira: projeções, previsões e técnicas*, L.R. Wong, R. Hakkert y R.A. Lima (comps.), São Paulo, Asociación Brasileña de Estudios Poblacionales (ABEP).