

NOTAS DE POBLACIÓN

AÑO XXVIII, N° 71, SANTIAGO DE CHILE



NACIONES UNIDAS

C E P A L

**Comisión Económica para América Latina y el Caribe
Centro Latinoamericano y Caribeño de Demografía (CELADE) – División de Población**

COMISIÓN ECONÓMICA PARA AMÉRICA LATINA Y EL CARIBE
José Antonio Ocampo Secretario Ejecutivo

CENTRO LATINOAMERICANO Y CARIBEÑO DE DEMOGRAFÍA
(CELADE) – DIVISIÓN DE POBLACIÓN
Daniel S. Blanchard Director

La Revista **NOTAS DE POBLACIÓN** es una publicación del Centro Latinoamericano y Caribeño de Demografía (CELADE) - División de Población, cuyo propósito principal es la difusión de investigaciones y estudios de población sobre América Latina y el Caribe, aun cuando recibe con particular interés artículos de especialistas de fuera de la región y, en algunos casos, contribuciones que se refieren a otras regiones del mundo. Se publica dos veces al año (junio y diciembre), con una orientación interdisciplinaria, por lo que acoge tantos artículos sobre demografía propiamente tal, como otros que aborden las relaciones entre las tendencias demográficas y los fenómenos económicos, sociales y biológicos.

Comité editorial:

Jorge Bravo
Rolando Sánchez
Susana Schkolnik

Coordinador técnico:

Juan Enrique Pemjean

Secretaría:

María Teresa Donoso

Redacción y administración:

Casilla 91, Santiago, Chile
E-mail: mdonoso@eclac.cl

Precio del ejemplar: US\$ 12

Suscripción anual: US\$ 20

Las opiniones expresadas en esta revista son responsabilidad de los autores, sin que el CELADE sea necesariamente partícipe de ellas.

SUMARIO

	<i>Página</i>
Presentación	7
Estimaciones indirectas de indicadores demográficos para áreas menores. Situación en América Latina. <i>Guiomar Bay</i>	9
Uso de variables sintomáticas para estimar la distribución espacial de la población. <i>María de Lourdes Teixeira Jardim</i> .	21
Variables sintomáticas en las estimaciones poblacionales a nivel cantonal en Costa Rica. <i>Edwin A. Chaves Esquivel</i>	51
Métodos de estimación para áreas pequeñas y una aplicación a la prevalencia anticonceptiva. <i>Alfredo Aliaga</i>	73
Estimación de la demanda de planificación familiar para los departamentos y provincias de Perú. <i>Guillermo Vallenás Ochoa</i>	91
Estimaciones de ingreso y pobreza para áreas geográficas menores: avances recientes en América Latina y el Caribe. <i>Jorge Bravo</i>	139
Combinación de datos censales y de encuestas para estudiar las dimensiones espaciales de la pobreza: el caso de Ecuador. <i>Jesko Hentschel, Jean Olson Lanjouw, Peter Lanjouw y Javier Poggi</i> .	149
Indicadores alternativos para focalizar el gasto social en Paraguay. <i>Marcos Robles</i>	185

PRESENTACIÓN

Los métodos de estimación indirecta de variables sociodemográficas para áreas geográficas menores han experimentado un franco desarrollo en América Latina y el Caribe. Aunque algunos de estos métodos tienen referentes en trabajos que datan de varias décadas atrás, durante los años noventa confluyeron tres factores principales, que dieron un gran impulso a esta línea de investigación: a) el fuerte aumento de la demanda de este tipo de información, originado en exigencias legales vinculadas a la representación política o a ciertas asignaciones presupuestarias, en el marco de procesos de descentralización, y en la focalización de las políticas sociales en general; b) una mayor frecuencia, periodicidad y sistematización del levantamiento de encuestas de hogares y demográficas con representación nacional, urbana y rural, y c) la más amplia disponibilidad de medios computacionales de bajo costo y alta velocidad de procesamiento de bases de datos con varios miles o millones de registros individuales, como las encuestas y los censos de población y vivienda.

En los artículos aquí reunidos, los temas y variables estudiados (población, prevalencia y demanda anticonceptiva, ingreso, pobreza), las fuentes de información utilizadas (registros vitales y administrativos, censos, encuestas) y los procedimientos de análisis y estimación (imputación proporcional simple, tipificación, análisis discriminante y de regresión) pueden aparecer, en una primera mirada, bastante disímiles. No obstante, los trabajos tienen importantes elementos en común, como el propósito de obtener estimaciones de indicadores de forma indirecta, cuando es imposible o insatisfactoria la estimación directa; el recurso a la combinación de dos o más fuentes de información, y la especificación y elaboración de modelos estadísticos, en general de tipo multivariado, apropiados para el objetivo que se persigue.

En los artículos quedan de manifiesto tanto la diversidad de indicadores y aproximaciones metodológicas de estimación como las principales dificultades que persisten en esta área. Sin embargo, lo que se destaca como aporte principal de esta colección son los grandes avances recientes en la superación de algunas limitaciones y la mejora de la calidad y facilidad de interpretación de los indicadores, así como de los análisis de política a que las estimaciones producidas pueden conducir.

La presente compilación forma parte de las actividades de un programa de cooperación técnica regional auspiciado conjuntamente por el Banco Interamericano de Desarrollo (BID) y la División de Población - Centro Latinoamericano y Caribeño de Demografía (CELADE) de la Comisión Económica para América Latina y el Caribe (CEPAL).¹ El volumen contiene también trabajos realizados en el Banco Mundial y por investigadores nacionales, en muchos casos efectuados en coordinación con los institutos nacionales de estadística. De hecho, un antecedente directo que sirvió de base para este volumen es el Taller sobre Estimaciones Sociodemográficas para Áreas Menores, celebrado en Lima entre el 27 y el 31 de marzo de 2000, evento organizado por el Instituto Nacional de Estadística e Informática (INEI) de Perú. El taller contó con el auspicio del Programa para el Mejoramiento de las Encuestas de Condiciones de Vida en América Latina y el Caribe (MECOVI), apoyado por la CEPAL, el BID y el Banco Mundial, y la asesoría técnica del programa BID/CELADE. Por tal motivo, el Comité Editorial de la revista agradece el apoyo de las instituciones mencionadas y las contribuciones sustantivas realizadas por los autores de los trabajos incluidos en este número especial de *Notas de Población*.

1 Programa de apoyo regional en el uso de información demográfica en la preparación, implementación y evaluación de proyectos sociales, ATN/TF-5827-RG.

**ESTIMACIONES INDIRECTAS DE INDICADORES
DEMOGRÁFICOS PARA ÁREAS MENORES.
SITUACIÓN EN AMÉRICA LATINA**

Guiomar Bay
CEPAL/CELADE - División de Población

RESUMEN

Se analizan las diferentes fuentes de información demográfica: censos de población, registro de estadísticas vitales y encuestas específicas, sus limitaciones, ventajas y desventajas para la estimación de indicadores demográficos a nivel de áreas menores. Teniendo en mente la dificultad de contar con la información necesaria para tales estimaciones, se analizan diferentes métodos de estimación "indirecta" de población y demanda de planificación familiar utilizados en América Latina, los cuales se basan en registros administrativos y en la combinación de fuentes. De los trabajos publicados en América Latina, pese las dificultades encontradas, se concluye que es factible utilizar tales métodos. Sin embargo, deben seguir los esfuerzos para mejorar las fuentes de información existentes y continuar investigando en el tema, así como probar tales métodos en otras realidades dentro de la región, en procura de tener más alternativas de estimación de indicadores demográficos para áreas menores.

INDIRECT ESTIMATION OF DEMOGRAPHIC INDICATORS FOR SMALL AREAS. THE SITUATION IN LATIN AMERICA

ABSTRACT

This article analyses such sources of demographic information as population censuses, vital records and specific surveys. It examines their limitations together with their advantages and disadvantages in estimating demographic indicators for small areas. In view of the difficulty of obtaining the information needed to prepare such estimates, a discussion is presented of various "indirect" methods for estimating population size and the demand for family planning services that are used in Latin America. These methods are based on the use of administrative records and on a combination of different sources. Although there have been some problems with the studies published in Latin America, they nonetheless indicate that it is feasible to employ these methods. Efforts to improve existing sources of information should continue, however, as should research on the subject. In addition, these methods should be tried out in other settings within the region so that a wider range of approaches to the estimation of demographic indicators for small areas can be developed.

1. INTRODUCCIÓN

Con el transcurso del tiempo han aumentado de manera creciente las necesidades de información desagregada geográficamente, para la ejecución de las estrategias de desarrollo vigente en los países de América Latina. Asimismo, las políticas y programas destinados a atender las necesidades básicas de la población dan mayor responsabilidad a los gobiernos locales, lo que se ha traducido en una mayor demanda de datos a ese nivel.

Por consiguiente, en la región creció la preocupación por obtener estimaciones de indicadores sociodemográficos a nivel cada vez más desagregado. Actualmente, las estimaciones de población para las áreas menores se elaboran básicamente por medio de extrapolaciones matemáticas de los niveles y tendencias observados en años recientes. No obstante, en la medida que las estimaciones se alejan del período base estos procedimientos suelen arrojar resultados distantes de la realidad, como consecuencia de los cambios en cuanto a la tendencia y nivel de las variables demográficas en períodos cortos, sobre todo de la migración. Por tanto, el uso de metodologías para determinar el número de habitantes basadas en tendencias anteriores es limitado, a menos que sean evaluadas y actualizadas periódicamente.

Por otro lado, si la necesidad es profundizar el análisis de las características específicas de la población para áreas menores, como por ejemplo la demanda de métodos anticonceptivos o demanda insatisfecha de planificación familiar, se presenta el problema de que las fuentes disponibles no consignan la información requerida al nivel geográfico que se desea estudiar. Así, surge la necesidad de combinar fuentes como, por ejemplo, encuestas y censos demográficos.

2. FUENTES DE INFORMACIÓN

Las principales fuentes de información utilizadas para la estimación de indicadores demográficos de áreas menores son los censos de población y vivienda, los registros de estadística vitales y las encuestas demográficas. En América Latina dichas fuentes de información suelen presentar problemas de cobertura, periodicidad y contenido, tanto en el ámbito nacional como en ámbitos más pequeños. Se han realizado esfuerzos para mejorar la calidad de la información, pero en general, éstos no han sido suficientemente eficaces. Muchas veces, este resultado se ha visto afectado por la situación socioeconómica o política de los países o áreas en cuestión.

Una forma de caracterizar cada fuente es según su cobertura, periodicidad y tipo de levantamiento (universo o muestra). Las principales características de los censos de población son su cobertura de carácter nacional, correspondiente a la información en un momento específico para toda la población, lo que permite el estudio de sus características a cualquier nivel geográfico. Los censos se realizan en general cada 10 años aunque en la región se observan muchas excepciones a esta norma.

Las encuestas demográficas tienen como característica principal el levantamiento de la información por muestreo. Generalmente se dedican a temas específicos, permiten investigar con mayor detalle las características estudiadas. Como son menores la escala y cobertura de las encuestas puede ejercerse mayor control en todas las etapas de recolección, procesamiento y divulgación de la información.

En principio la cobertura de los registros de estadísticas vitales es nacional. La mantención de los registros es un proceso continuo de recolección de datos sobre toda la población que permite obtener, en teoría, información para todos los niveles geográficos. En muchos países de la región, no obstante, la cobertura real es incompleta.

Entre las limitaciones que presentan las diversas fuentes pueden considerarse dos tipos de deficiencias: errores de cobertura y de contenido. El error de cobertura se refiere a la omisión sobre ciertos temas de un cuestionario, de personas de una vivienda u hogar, viviendas o áreas completas, e incluso la duplicidad de la información. Esa duplicidad puede ocurrir principalmente en los censos de población cuando el trabajo de campo dura varios días, a veces semanas dependiendo del país, entre otras causas. En el caso de los registros de estadísticas vitales también se presentan problemas de registro tardío.

Por error de contenido se entienden las deficiencias de la declaración de la información, ya sea porque la pregunta fue mal interpretada por el informante o el empadronador o el informante declaró los datos de otra persona, o simplemente se ignoraba la respuesta correcta.

Los errores que presentan las diversas fuentes pueden originarse en cualquier etapa del proceso de recolección de la información. En las etapas previas (de planeación y de organización) pueden surgir problemas en la actualización de la cartografía, el diseño del cuestionario o producirse recortes presupuestarios, etc. En el trabajo de campo se presentan problemas de comunicación entre el informante y el empadronador ya sea por razones conceptuales; mala interpretación de la pregunta o de la respuesta; errores en el registro de la información; desconocimiento de la respuesta correcta, o errores deliberados. Por último cabe mencionar las equivocaciones que se pueden producir al momento de procesar y divulgar la información o esos errores pueden variar según el área geográfica.

Lo expuesto anteriormente permite fácilmente visualizar las ventajas de una u otra fuente. Los censos de población, por ejemplo, tienen la gran ventaja de su cobertura a escala nacional y un levantamiento de la información dirigido a todo el universo. Por otro lado, tienen la desventaja de ser un proceso de gran envergadura y, por consiguiente, son difíciles de planificar y organizar, lo que no permite realizarlos con una frecuencia mayor y tienen limitaciones implícitas en cuanto a la profundidad de la información recogida sobre los temas en estudio.

Las encuestas demográficas, dada su característica de levantamiento por muestreo, son procesos de menor envergadura y así permiten un mayor control de todas sus etapas y profundizar más los temas investigados. La gran limitación de las encuestas es que sólo permiten estudiar las características de áreas geográficas según el nivel de representatividad de la muestra.

Los registros de estadísticas vitales presentan principalmente problemas de cobertura, aunque en teoría deberían abarcar a toda la población. En América Latina son pocos los países que poseen estadísticas vitales confiables en el ámbito nacional y local.

3. ESTIMACIÓN DE INDICADORES DEMOGRÁFICOS PARA ÁREAS MENORES

3.1 Estimaciones de población para áreas pequeñas

La necesidad cada vez mayor de contar con información sobre áreas menores ha llevado a desarrollar métodos que consideran las fuentes de información disponibles y sus limitaciones. Cuando el objetivo es estimar la población por áreas en períodos intercensales surge la dificultad de estimar la población migrante entre áreas pequeñas. Además, no siempre se pueden estimar las demás variables de la dinámica demográfica (fecundidad y mortalidad) al nivel de áreas menores porque a ese nivel no hay información o su calidad es deficiente. Eso impide aplicar el método clásico de estimación de la población, cual es el de la ecuación compensadora. Además, a lo anterior se suma en muchos casos el problema del desmembramiento y/o reagrupación de áreas a raíz de los cuales se crean nuevas unidades político-administrativas.

Al respecto, una alternativa útil para determinar el número de habitantes en áreas pequeñas es la aplicación de métodos basados en indicadores indirectos del tamaño de la población, como los que se basan en variables sintomáticas. Esas metodologías han sido utilizadas en los Estados Unidos, Canadá y, en nuestra región, en un Estado de Brasil, en Chile y Costa Rica (véase Chaves y Teixeira en este volumen, y Bay, 1998)

Otro aspecto pertinente que debe considerarse para la utilización de estos métodos es su fácil aplicación, ya que permiten, por un lado, la actualización de la información sobre población en áreas pequeñas y, por otro lado, detectar cambios en la tendencia del movimiento y crecimiento intercensal de la población. Asimismo, cabe resaltar que también podrían obtenerse estimaciones de población para nuevas divisiones administrativas, aunque éstas no hubieran sido consideradas en la elaboración de los modelos.

Los métodos ejemplificados en Bay (1998), Chaves y Teixeira (*op.cit.*) —unos sencillos y otros más sofisticados— pueden aplicarse según la cantidad de información disponible y su calidad. Tales métodos requieren registros periódicos de variables como nacimientos, defunciones, registro de automóviles, impuestos, construcción de viviendas, matrícula escolar, etc. En muchos países de América Latina, estos registros suelen tener problemas de omisión diferencial por área,

interrupción del registro, alteraciones frecuentes en la división político-administrativa, cambios en los sistemas económicos, programas sociales de vivienda, educación, salud, servicios básicos. A veces, la información disponible simplemente no se tabula ni publica. Sin embargo, en algunos casos, conociendo tales limitaciones pueden introducirse, algunas modificaciones a los procesos de estos registros como, por ejemplo, tabular los datos según área político-administrativa, que permiten obtener la información necesaria para la aplicación de dichos métodos.

3.1.1 Información básica

Como un primer paso es necesario investigar la mayor cantidad de información disponible acerca de las variables sintomáticas. A continuación es necesario verificar la disponibilidad periódica de cada una de esas variables sintomáticas, su calidad diferencial por área, las posibles interrupciones o creación de nuevos registros, los cambios estacionales importantes, los incentivos especiales aplicados a ciertos registros, las modificaciones de las leyes o definiciones del hecho registrado, etc.

Al respecto, cabe destacar la enorme importancia que tienen el “contacto” con la información y el análisis crítico de los sistemas de recolección, a fin de lograr una profunda familiarización con la información que se ha de utilizar. Gracias a eso se puede detectar, por ejemplo, posibles errores de estimación atribuidos a problemas de la calidad de información diferencial por área, o a la compatibilización entre períodos, entre otros. Sin duda, buena parte del tiempo y esfuerzo empleado en la investigación debe orientarse a la evaluación exhaustiva de la información, pues de ella dependerá la adecuada elección del modelo de estimación que se usará.

Dados los pasos anteriormente descritos, se procede a llenar lo que se podría llamar el inventario de la información disponible, con el cual se obtiene una lista de variables por área y períodos disponibles, así como también la población de las mismas áreas, en los últimos dos censos, procediendo a elegir los métodos factibles de aplicarse según la realidad de cada país.

3.1.2 Evaluación general de los diferentes métodos de estimación utilizando variables sintomáticas

Si bien es bastante limitada la disponibilidad de variables sintomáticas que permitan aplicar los diferentes métodos de estimación de áreas menores en América Latina, fue posible encontrar modelos que suministraran estimaciones razonables en los tres estudios comentados en este artículo.

Cabe destacar que, entre los métodos aplicados, el de correlación de razón fue el que produjo las estimaciones más precisas. La mayor imprecisión de los métodos más sencillos, distribución por prorrateo o proporcional, está directamente relacionada con las diferencias de cobertura o calidad de ciertos datos entre unidades territoriales, diferencias que tienden a ser menores cuando se consideran las variaciones de la población y las variables sintomáticas en el tiempo.

Como cabía esperar, los mayores problemas de precisión de las estimaciones se producen en aquellas áreas en que es menor el tamaño de la población. En vista de esto, la estratificación de las áreas según tamaño resulta de gran importancia pues mejora la precisión de las estimaciones. Una alternativa consiste en agrupar algunas de estas áreas en el momento de estimar el modelo y/o la población, utilizando posteriormente algún criterio para desagregarlas.

En resumen, es factible utilizar métodos que se basan en variables sintomáticas para estimar el volumen de la población de áreas menores en América Latina. La identificación del modelo adecuado dependerá de la calidad y periodicidad de las variables así como de un cuidadoso escrutinio de los diferentes métodos por parte de los investigadores.

3.2 Estimación de la demanda de planificación familiar para áreas menores

Así como ha aumentado la demanda de información de población a nivel más desagregado, también ha crecido la de información sobre otras características sociodemográficas, como por ejemplo, la demanda de planificación familiar, con el propósito de verificar la cobertura y calidad nacional de determinados programas, o para la implementación de programas.

Cuando se desea estimar estas características, se presenta el problema de la falta de información desagregada según áreas

geográficas menores. El problema surge de la dificultad de investigar tales características en los censos de población de cobertura nacional, y de la imposibilidad de estimarlos por medio de encuestas específicas, cuyo diseño muestral no contempla la representatividad al nivel de desagregación requerido.

Este problema se aborda en los artículos de Aliaga y Vallenas (en este volumen) sobre los métodos de estimación de la prevalencia anticonceptiva y la demanda insatisfecha de planificación familiar. Un enfoque similar ha sido usado en los estudios de Robles y Hentschel y otros (en este volumen) para estimar ingreso y pobreza para áreas geográficas menores. La dificultad fundamental es que encuestas específicas, principal fuente disponible para estudiar estos indicadores, no prevén la representatividad del diseño muestral al nivel requerido, lo que hace que las estimaciones directas presenten errores muestrales de gran magnitud.

La alternativa propuesta por los autores son las estimaciones indirectas. En ese proceso de estimación se supone que los indicadores que se desea estimar a nivel de áreas menores tienen un comportamiento similar a los del área mayor que las contiene, con respecto a un conjunto de variables explicativas. Para aplicar esta metodología es necesario combinar dos fuentes (encuesta y censo) a fin de estimar estos indicadores a nivel del área menor por medio de un modelo de regresión. Se utilizan los datos de una encuesta para estimar el modelo, el que se aplica al nivel geográfico menor usando los datos censales. Los modelos se construyen con las variables comunes en las dos fuentes que estén correlacionadas con el indicador que se desea estimar.

Los resultados para la demanda de planificación familiar en República Dominicana y Kenia (Aliaga), así como en Perú (Vallenas) son bastante satisfactorios, y constituyen una buena base para desarrollar este tipo de estimaciones en el futuro en éstos u otros países de la región.

4. CONCLUSIONES Y RECOMENDACIONES

1. La necesidad de contar con mejores estimaciones de población al nivel de áreas menores (localidades, municipios, distritos, etc.) está dada por el creciente peso de la comunidad en lo económico, político, y cultural. Sea para la elaboración de planes del gobierno central, por

iniciativas que surgen de la propia comunidad, o por un deliberado impulso a la descentralización y al desarrollo local, lo cierto es que cada vez será más necesario y útil disponer de cifras precisas y actualizadas sobre el tamaño de la población de áreas menores, sus características y necesidades.

2. Si se dispusiera de registros frecuentes del movimiento migratorio entre estas áreas y de un recuento completo de los nacimientos y de las defunciones, estaría resuelto el problema de la actualización del tamaño de la población, pues estarían determinados los tres componentes de la dinámica demográfica. En los numerosos casos donde no se cumplen las condiciones de este escenario ideal es muy importante contar con métodos alternativos que indiquen la variación de la población de áreas menores.

3. Censar la población más frecuentemente sería otra alternativa, pero es poco factible por los altos costos que implica esta operación. Cabe recordar que a pesar de la recomendación de las Naciones Unidas de que se realicen los censos cada 10 años, muchos países de América Latina los llevan a cabo a intervalos superiores.

4. Teniendo en cuenta las razones expuestas, se cree que una alternativa eficiente y eficaz sería mejorar los procedimientos de registros administrativos ya existentes, para su posterior utilización en actualización de las estimaciones de población. En los trabajos realizados en la región, se demuestra que, a pesar de los problemas existentes en cuanto a disponibilidad y calidad de información, se puede llegar a estimaciones bastante razonables de la población de las áreas consideradas.

5. La selección del modelo o método de estimación de indicadores socioeconómicos no es tarea fácil y dependerá de la calidad de la información y su disponibilidad, y también de las características particulares de las áreas para las cuales se elaborarán las estimaciones. Se considera bastante útil elaborar un estudio previo, evaluando a priori de la situación en cada caso concreto, y familiarizarse con la información y los distintos métodos, a fin de escoger el más adecuado. Además, resultará muy conveniente realizar evaluaciones a posteriori, lo que permitirá introducir cambios en el modelo utilizado.

6. Otro aspecto importante que conviene destacar es el control permanente de la información básica, o sea, de los indicadores utilizados, lo que podría llevar incluso al cambiar el procedimiento de estimación.

7. Se hace necesaria una mayor integración de las diversas fuentes de información y su accesibilidad por parte del organismo responsable de

las estimaciones de población. A veces, introducir pequeñas modificaciones en los procesos de recolección de información o tabulación de la misma, o en ambos, abre una gama de posibilidades de construcción de modelos, sobre todo de criterios para la estratificación de las áreas a estimar. De esta forma se podría disponer de un sistema confiable —y de bajo costo— para la realización de estimaciones de población.

8. Se puede concluir que es factible obtener estimaciones de indicadores sociodemográficos para áreas menores en América Latina utilizando las fuentes de información existentes. Por ello basándose en las experiencias ya realizadas, es útil apoyar las actividades que apunten a una mayor discusión y divulgación de los trabajos realizados, que es la línea en la que se inscribe la presente compilación de estudios sobre esta temática.

BIBLIOGRAFÍA

- Aliaga, Alfredo (2000), “Developments in small area estimation of contraceptive prevalence and mayor issues confronted”, *Notas de población* N° 71, Santiago de Chile, Comisión Económica para América Latina y el Caribe, Centro Latinoamericano y Caribeño de Demografía - División de Población (CEPAL/CELADE).
- Bay, Guiomar (1998), “El uso de variables sintomáticas en la estimación de la población de áreas menores”, *Notas de Población*, N° 67/68 (LC/G.2048; LC/DEM/G.186), Santiago de Chile, Comisión Económica para América Latina y el Caribe, Centro Latinoamericano y Caribeño de Demografía - División de Población (CEPAL/CELADE).
- Chackiel, Juan y Macció, Guillermo (1978), “Evaluación y corrección de datos demográficos”, Capítulos I y II, serie B, N° 39, Santiago de Chile.
- Chaves, Edwin A. (2000), “Variables sintomáticas en las estimaciones de población a nivel cantonal en Costa Rica”, *Notas de población* No 71, Santiago de Chile, Comisión Económica para América Latina y el Caribe, Centro Latinoamericano y Caribeño de Demografía - División de Población (CEPAL/CELADE).
- Florez Carmen Elisa (1989), “Las proyecciones de población y la estructura económica y social”, documento presentado en el

- Seminario Internacional sobre Proyecciones Subnacionales de Población*, Girardot, 31 de octubre a 2 de noviembre de 1988.
- Rueda, José Olinto (1989), "Las proyecciones subnacionales de población y la planificación del desarrollo", Documento presentado en el *Seminario Internacional sobre Proyecciones Subnacionales de Población*, Girardot, 31 de octubre a 2 de noviembre de 1988.
- Teixeira Jardim, Maria de Lourdes (2000), "El uso de variables sintomáticas para estimar la distribución espacial de población: aplicación a los municipios de Río Grande del Sur, Brasil", *Notas de población* N° 71, Santiago de Chile, Comisión Económica para América Latina y el Caribe, Centro Latinoamericano y Caribeño de Demografía - División de Población (CEPAL/CELADE).
- Vallenas Ochoa, Guillermo (2000), "Perú: demanda de planificación familiar, en áreas pequeñas", *Notas de población* N° 71, Santiago de Chile, Comisión Económica para América Latina y el Caribe, Centro Latinoamericano y Caribeño de Demografía - División de Población (CEPAL/CELADE).

**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 prorateo; 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 prorrateo 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).

**VARIABLES SINTOMÁTICAS EN LAS ESTIMACIONES
POBLACIONALES A NIVEL CANTONAL
EN COSTA RICA**

Edwin A. Chaves Esquivel
Universidad de Costa Rica
Universidad Nacional Autónoma de Costa Rica

RESUMEN

Dado que Costa Rica no realizó el censo de población y vivienda correspondiente a la ronda de 1990, ha sido necesario encontrar técnicas que permitan estimar sus principales características demográficas. En este artículo se analiza el uso de variables tales como el padrón electoral, los datos vitales y la matrícula escolar en la realización de estimaciones poblacionales a nivel cantonal. Para ello se aplicaron cuatro métodos que demostraron ser buenas alternativas.

**SYMPTOMATIC VARIABLES IN POPULATION
ESTIMATES AT THE CANTON LEVEL
IN COSTA RICA**

ABSTRACT

Since Costa Rica did not carry out a population and housing census for the 1990 round, it has been necessary to find other methods for arriving at estimates of the country's main demographic features. This article analyses the use of such variables as voting lists, vital statistics and school enrolment records in the preparation of population estimates at the canton level. To this end, four methodologies that proved to be satisfactory options were applied.

INTRODUCCIÓN

Debido a la problemática que conllevan las estimaciones poblacionales a nivel de áreas administrativas menores, desde hace muchas décadas diversos investigadores han propuesto utilizar variables sintomáticas. Éstas son variables auxiliares ligadas al crecimiento poblacional, que pueden ser utilizadas para medir sus cambios, y que se construyen a partir de información recolectada regularmente por empresas e instituciones públicas o privadas con fines de orden administrativo; se trata, entre otras, de la matrícula escolar, las declaraciones de impuestos, el número de abonados a los servicios básicos y de electores inscritos en el padrón electoral (Serow y Rives, 1995; Simpson y otros, 1996).

La intención de recolectar estos datos no es estimar poblaciones. Esto obliga a estar pendiente de ciertos problemas como el cambio de las definiciones en el tiempo y las interrupciones en la recopilación, que pueden restringir las posibilidades de aplicación con fines estimativos.

Las técnicas utilizadas para trabajar con estas variables son muy variadas. Evidentemente, la precisión de las estimaciones depende de la relación existente entre las variables sintomáticas y el tamaño de la población, así como del grado de relación a lo largo del tiempo, pero también influye la disponibilidad de datos con el detalle debido.

El presente estudio ofrece una contribución en este sentido. Se determinan las variables sintomáticas más estrechamente asociadas al crecimiento poblacional, y se aplican cuatro métodos de estimación que han dado resultados aceptables en países desarrollados. Asimismo, se efectúan estimaciones poblacionales a nivel cantonal para 1984 (año del último censo), que luego se comparan con su valor censal, para medir la precisión de cada uno de los indicadores.

I. ESTRATEGIA METODOLÓGICA

El presente estudio se basa en la aplicación de cuatro métodos que, por las características del medio y de la información disponible, pueden ofrecer aportes importantes en esta materia, a nivel nacional. Estos métodos permiten hacer estimaciones para las que se requiere conocer el tamaño de la población de un área mayor que comprenda una serie de áreas más pequeñas. Por tal razón, las estimaciones deben prorratearse, de manera que el conjunto de ellas reproduzca la población total del área.

1. Método de razón censal

En este método se consideran las tasas de ocurrencia de un determinado indicador sintomático, partiendo del supuesto de que las áreas locales presentan una razón de cambio proporcional a la del área mayor, en el período comprendido entre el último censo y la fecha de estimación (Voss y otros, 1995).

En primer lugar, el procedimiento consiste en determinar la tasa de ocurrencia de cada variable sintomática para las distintas áreas menores, durante el último período censal. Si se desea estimar la población a partir de la variable sintomática S para un cantón u , t años después del último censo, se utiliza la siguiente ecuación:

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

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

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

$P_0(u)$ es la población del cantón u medida en el último censo.

Posteriormente, se aproxima la tasa de la variable sintomática a la fecha t , por medio de un parámetro ϕ , que representa la razón de cambio en la variable sintomática entre el año censal y la fecha t , de tal modo que:

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

Para estimar el valor de ϕ es necesario tener información del área mayor que abarque a u , sobre la cual se pueden hacer estimaciones independientes de la población en la fecha t . Esta área se designa como M y su población en la fecha t como $P_t(M)$, donde la estimación de la población de M en t es $\hat{P}_t(M)$. De este modo, se calculan las tasas de ocurrencia de S en el área M , en la fecha del censo y la fecha t , utilizando la fórmula [1]. Si la razón de cambio en la variable sintomática para el área mayor M puede aproximarse efectivamente al valor de ϕ , se tiene:

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

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

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

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

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

Generalmente, se obtiene más de una estimación de $P_t(u)$ al considerar diferentes variables sintomáticas. La estimación final es, por lo general, el promedio simple de las estimaciones individuales, aunque hay métodos para ponderarlas (Ericksen, 1973; Ericksen, 1974).

La principal ventaja de este método es la simplicidad de su aplicación; su desventaja es que la dirección de las series sintomáticas podría verse 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 podría, incluso,

evolucionar en dirección contraria a la razón del área mayor; en tal caso, no se cumpliría el supuesto básico del modelo.

2. 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 de la variable sintomática para cada cantón. De acuerdo a este método se supone que el crecimiento de esta tasa para el cantón u es igual al crecimiento registrado en el área mayor M .

Considerando nuevamente el cálculo de estimaciones por medio de la variable sintomática S , la aproximación para la tasa de ocurrencia de este indicador en la fecha t , para el cantón u , viene dada por la siguiente fórmula:

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

donde sus componentes se definen como se especificó anteriormente.

3. Método de correlación de razón

La aplicación de este método consiste fundamentalmente en dividir la población de un área mayor en los cantones que la constituyen. Esta técnica ha demostrado ser una de las más exactas en la estimación de población poscensal (Namboodiri, 1972; Mandell y Tayman, 1982; Fenney, Hibbs y Gillaspy, 1995).

El método se basa en el supuesto de que existe una relación lineal entre las razones de cambio de las proporciones poblacionales y las razones de cambio de las proporciones de las variables sintomáticas entre dos períodos. Considerando, además, que los cambios en esta relación a lo largo del tiempo afectan uniformemente todos los cantones, se desarrolla un modelo más refinado, que puede formularse de la siguiente manera:

$$Y_u = a_0 + a_1 X_u \quad [7]$$

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

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 las fechas en que se realizaron el penúltimo y el último censo, respectivamente.

Esta ecuación indica que la razón de cambio a lo largo del tiempo de la proporción poblacional está sistemáticamente relacionada con la correspondiente razón de cambio en las proporciones de los indicadores. Se pueden introducir en el modelo otras variables sintomáticas, de tal 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 variable sintomática j .

Una vez estimado el modelo de regresión, es muy sencillo hacer las estimaciones de población para un período poscensal.

4. Método de correlación de tasa

Se le critica a la técnica de correlación de razón una inconsistencia interna, ya que los coeficientes del modelo se calculan tomando como referencia el tiempo transcurrido entre los censos. Sin embargo, las estimaciones no coinciden con el período base, que por lo general, es menor que el tiempo transcurrido 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 modificación del método de correlación de razón llamada “tasa de correlación”, que se basa en una aproximación exponencial de cambio y se realiza tomando el logaritmo natural de cada razón y dividiéndolo por el tiempo transcurrido entre los censos utilizados como referencia (Swanson y Tedrow, 1984). Formalmente, el modelo se puede expresar de la siguiente manera:

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

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

Y_u y $X_{u,j}$ están definidos tal como se indicó en el caso de la técnica de correlación de razón y k es el intervalo de tiempo transcurrido entre los censos.

Después de un experimento que se llevó a cabo en varios condados del estado de Washington, Swanson y Katzoff señalan tres razones para suponer que la técnica de correlación de tasa da resultados más precisos, que la de correlación de razón:

i) Como la primera proviene de una transformación de la segunda, se deduce que ambas tienen la misma tendencia. Si ésta se determina durante varios años de estimaciones, el método de correlación de razón presentará mayores inconsistencias, porque estas extrapolaciones están basadas en una relación estructural de cambio a lo largo de un período mayor de años.

ii) La transformación de las tasas reduce las variancias, por lo que se podría lograr una mayor exactitud con la correlación de tasa.

iii) Puesto que se ha reducido el efecto de la autocorrelación espacial, es de esperar que las estimaciones sean más precisas (Swanson y Tedrow, 1984).

5. Variables sintomáticas disponibles

Para cumplir con los objetivos del estudio es requisito fundamental contar con información sintomática sobre las diferentes áreas

administrativas del país. Debido a las características de los métodos de estimación empleados, era necesario que los datos utilizados estuvieran disponibles a partir del censo de población de 1963. Se seleccionaron siete fuentes de datos: matrícula escolar en primer y segundo ciclos, proporcionada por el Ministerio de Educación Pública; nacimientos; defunciones; permisos de construcción de viviendas, suministrados por la Dirección General de Estadística y Censos; registros de asegurados de la Caja Costarricense del Seguro Social; registros de atención en los centros de salud facilitados por el Ministerio de Salud y número de electores inscritos en el padrón electoral, proporcionado por el Tribunal Supremo de Elecciones.

Cada una de estas fuentes de datos debió ser sometida a un proceso de revisión y evaluación, con el fin de evitar inconsistencias (Chaves, 1997).

6. Valoración de las estimaciones

La mejor manera de evaluar la precisión de las estimaciones producidas por los diferentes métodos, es compararlas con sus valores reales. Sin embargo, debido a que este valor es desconocido, la comparación se hace con los valores censales. Para ello, fue necesario conseguir estimaciones poblacionales correspondientes a 1984, año del último censo poblacional. Éstas se realizaron aplicando los diferentes métodos de estimación, con todas las variables sintomáticas seleccionadas y combinaciones de éstas.

Para medir la calidad de las estimaciones producidas se calculó el error porcentual (*EP*) o porcentaje de desviación respecto del valor censal. La medición se realiza aplicando la siguiente fórmula:

$$EP = \frac{(\text{Población estimada} - \text{Población censal})}{\text{Población censal}} \times 100 \quad [10]$$

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

Puesto que las estimaciones se prorratan para ajustar la población total del país, el error porcentual promedio presenta valores cercanos a cero. Por tal razón, en el análisis del comportamiento global de las

estimaciones fue necesario calcular el valor absoluto del error porcentual (*AEP*), utilizando la siguiente ecuación:

$$AEP = \left| \frac{\text{Población estimada} - \text{Población censal}}{\text{Población censal}} \right| \times 100 \quad [11]$$

De este modo, el promedio de *AEP* resultó ser adecuado para evaluar la calidad de un conjunto de estimaciones y un buen parámetro de comparación de los resultados de dos o más conjuntos.

Por otra parte, se analizó el comportamiento de *AEP* con respecto a la magnitud de la población del cantón considerado. Para medir el efecto que tiene dicha magnitud, se calculó el promedio de *AEP* ponderado por la población censal:

$$\frac{\sum_j (AEP)_j \cdot P_j}{\sum_j P_j} \quad [12]$$

donde P_j es la población censal del cantón j .

Si el tamaño del cantón no afecta las estimaciones, dicho promedio no debió ser muy diferente del promedio simple de *AEP*.

7. Precisión de las estimaciones

De acuerdo con lo anterior, la precisión de las estimaciones se determinó mediante una comparación de su valor con la información censal. Para calificar una estimación como adecuada o inadecuada, se consideró que la información censal está sujeta a errores de cobertura. En el último censo, este error fue estimado aproximadamente en un 6% (CELADE/DGEC/MIDEPLAN, 1988). En consecuencia, no parece lógico que se les exija una extrema precisión a las estimaciones poblacionales de áreas menores. Por esta razón, para efectos del presente trabajo se consideraron “adecuadas” aquellas estimaciones que presentaron errores menores al 5%, “aceptables” las que estuvieron entre 5% y 10% e “inadecuadas” las que superaron el 10%.

II. RESULTADOS

1. Precisión y aplicación simple de las variables sintomáticas

A continuación se analiza el comportamiento de cada una de las variables, consideradas en el estudio en el período 1973-1984. En una primera etapa, se seleccionaron 22 cantones de las provincias de San José y Heredia, con respecto a cada uno de los cuales se recopiló información correspondiente a las siete variables consideradas en los últimos dos años censales (1973 y 1984). Se determinó, tanto para la población como para cada variable, la razón de cambio con respecto a la proporción que representa cada cantón en el total del país, en el período 1973-1984.

En el cuadro 1 se presenta la matriz de correlación de estas razones de cambio. Los mayores coeficientes de correlación con la población corresponden a los indicadores “padrón electoral” y “nacimientos”, y son superiores a 0.80. Las defunciones y la matrícula escolar también presentaron un coeficiente de correlación superior a 0.60, por lo que es importante considerar su aporte al proceso de estimación. Las otras tres variables presentan un coeficiente de correlación muy bajo, por lo que se decidió excluirlas del estudio.

Las variables sintomáticas seleccionadas se utilizaron en la segunda etapa con el objeto de hacer estimaciones poblacionales para todos los cantones del país.

Como ya se ha mencionado, para aplicar los métodos de correlación de razón y correlación de tasa se requiere información de los dos censos anteriores. El cálculo de las estimaciones poblacionales para 1984 exige utilizar los censos de 1963 y 1973, lo que obligó a trabajar con los 68 cantones existentes en el país en 1963. Por esta razón, la información correspondiente a los cantones creados después de este año se incluyó en el lugar de procedencia de cada uno. Este problema no afecta a los métodos de razón censal y diferencia de tasas, que son los únicos para los cuales se requiere información del censo de 1973, cuando había 79 cantones. Sin embargo, para efectos comparativos, se decidió considerar 68 en todos los casos.

Por otra parte, para dar mayor estabilidad a la estimación de los nacimientos y defunciones, se decidió calcular un promedio trienal, en lugar del valor simple, lo que reduce la variabilidad de estas variables sintomáticas.

Cuadro 1
**MATRIZ DE CORRELACIÓN DE LAS RAZONES DE CAMBIO -
 EN LAS PROPORCIONES DEL TAMAÑO POBLACIONAL Y DE OCHO
 VARIABLES SINTOMÁTICAS EN 22 CANTONES DE SAN JOSÉ Y HEREDIA
 1973-1984**

	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
Población	1.00							
Padrón electoral	0.88	1.00						
Nacimientos	0.87	0.79	1.00					
Defunciones	0.65	0.58	0.56	1.00				
Matrícula escolar	0.61	0.36	0.48	0.21	1.00			
Asistencia centro de salud	0.35	0.38	0.10	0.47	0.12	1.00		
Permisos de construcción	0.23	0.07	0.00	0.04	-0.02	0.30	1.00	
Afiliación a la CCSS	-0.31	-0.18	-0.27	-0.23	-0.48	-0.07	0.09	1.00

Fuente: Elaboración propia sobre la base de información proporcionada por la Dirección General de Estadística y Censos (DGEC), Caja Costarricense de Seguro Social (CCSS), Ministerio de Educación Pública, Ministerio de Salud, Tribunal Supremo de Elecciones.

(a): Población; (b): padrón electoral; (c): nacimientos; (d): defunciones; (e): matrícula escolar; (f): asistencia a centro de salud; (g): permisos de construcción; (h): afiliación a la CCSS.

En el cuadro 2 se presentan los errores producidos en las estimaciones, para las cuatro variables sintomáticas consideradas.

Al igual que en la primera etapa, el padrón electoral, los nacimientos y el crecimiento vegetativo son las variables que dan los mejores resultados a este nivel. El padrón electoral es evidentemente superior a los demás indicadores. Al utilizar el padrón electoral en las estimaciones poblacionales, el error de estimación no superó el 5%, y ninguna de las demás variables sintomáticas se aproximó a este valor. Los nacimientos y el crecimiento vegetativo también ofrecen buenos resultados. Las estimaciones derivadas de la matrícula escolar y de las defunciones contienen grandes errores, en una proporción muy alta errores porcentuales que superan el 10%, por lo que estas estimaciones deben utilizarse con gran cautela.

Cuadro 2
**COMPARACIÓN DEL VALOR ABSOLUTO DEL ERROR PORCENTUAL
 DE (AEP) LAS ESTIMACIONES POBLACIONALES A NIVEL
 CANTONAL POR VARIABLE SINTOMÁTICA
 SEGÚN MÉTODO UTILIZADO, 1984**

Método	Variable sintomática			
	Padrón electoral	Naci- mientos	Matrícula escolar	Defun- ciones
Razón censal				
Promedio de AEP	5.0	8.8	11.3	15.9
Porcentaje de AEP inferior a 5%	61.8	42.6	30.9	14.7
Porcentaje de AEP superior a 10%	13.2	33.8	51.5	67.6
Diferencia de tasas				
Promedio de AEP	4.3	8.3	13.1	17.9
Porcentaje de AEP inferior a 5%	66.2	44.1	22.1	16.2
Porcentaje de AEP superior a 10%	5.9	32.4	58.8	72.1
Razón correlación				
Promedio de AEP	5.0	6.7	9.5	10.5
Porcentaje de AEP inferior a 5%	58.8	51.5	32.4	29.4
Porcentaje de AEP superior a 10%	13.2	20.6	41.2	36.8
Tasa correlación				
Promedio de AEP	4.9	6.4	9.6	10.4
Porcentaje de AEP inferior a 5%	55.9	51.5	36.8	27.9
Porcentaje de AEP superior a 10%	11.8	23.5	42.6	44.1

Fuente: Elaboración propia.

2. Aplicación de métodos de estimación con múltiples variables

A continuación se presentan los resultados obtenidos con más de una variable sintomática. En los casos de la razón censal y la diferencia de tasas, las estimaciones basadas en la combinación de dos o más variables corresponden al promedio de estimaciones simples. En los de correlación de razón y correlación de tasa, basta con incluir las variables pertinentes en el modelo de regresión múltiple.

En el cuadro 3 se indica la combinación de las variables que mostraron la mayor precisión en las estimaciones.

Cuando se utiliza el padrón electoral en combinación con los nacimientos como variables sintomáticas se obtienen buenos resultados, fundamentalmente cuando se emplean métodos basados en modelos de regresión. De este modo, el 72% de los cantones presentan un *AEP* inferior al 5%, un promedio de *AEP* de aproximadamente un 4% y una desviación estándar de los *EP* cercana al 5%. La combinación de la matrícula escolar, el padrón electoral y los nacimientos, reduce el promedio de *AEP* en 0.3 puntos porcentuales; en el caso de la correlación de tasa, disminuye el número de cantones con *AEP* superior al 10%.

Los resultados revelan cierta influencia de la magnitud de la población del área en las estimaciones, pues en todos los modelos se reduce el error promedio cuando se pondera por esta variable.

En el cuadro 4 se presentan los valores de R^2 correspondientes a los cuatro modelos del cuadro 3 y la significación de los coeficientes de cada variable sintomática, calculados mediante la prueba estadística “*t*”.

Todos los modelos empleados muestran un R^2 alto y los indicadores utilizados son significativos al nivel del 5%, como lo muestra la prueba “*t*”, con la excepción de la matrícula escolar. Este resultado señala que el aporte de la matrícula escolar al modelo que incluye el padrón electoral y los nacimientos es mínimo, dado que R^2 únicamente aumenta en un punto porcentual.

Con respecto al cumplimiento de los supuestos del modelo de regresión, las pruebas hechas demuestran que el modelo de correlación de tasa ofrece una mayor consistencia, debido fundamentalmente a que el método de correlación de razón presenta problemas con el supuesto de variancia constante (Chaves, 1997).

Cuadro 3

**COMPARACIÓN DEL VALOR ABSOLUTO DEL ERROR PORCENTUAL (AEP)
Y DEL ERROR PORCENTUAL (EP) EN LAS ESTIMACIONES
POBLACIONALES A NIVEL CANTONAL POR VARIABLE
SINTOMÁTICA SEGÚN MÉTODO UTILIZADO, 1984**

Variable sintomática utilizada	Método			
	Razón censal	Diferencia de tasas	Correlación de razón	Corre- lación de tasa
Padrón electoral				
Promedio de AEP	5.0	4.3	5.0	4.9
Desviación estándar de los EP	6.5	5.5	6.4	6.3
Promedio ponderado de AEP	4.8	4.3	4.6	4.6
Valor máximo de AEP	21.2	16.6	18.5	19.4
Porcentaje de AEP inferior a 5%	61.8	66.2	58.8	55.9
Porcentaje de AEP superior a 10%	11.8	5.9	13.2	11.8
Padrón electoral y nacimientos				
Promedio de AEP	4.7	4.7	4.1	4.0
Desviación estándar de los EP	6.8	6.1	4.9	4.8
Promedio ponderado de AEP	3.8	4.0	3.9	3.9
Valor máximo de AEP	21.6	20.7	12.1	12.5
Porcentaje de AEP inferior a 5%	66.2	64.7	72.1	72.1
Porcentaje de AEP superior a 10%	10.3	10.3	2.9	2.9
Padrón electoral, nacimien- tos y matrícula escolar				
Promedio de AEP	4.8	5.4	3.8	3.8
Desviación estándar de los EP	5.9	6.6	4.7	4.6
Promedio ponderado de AEP	4.1	4.8	3.7	3.6
Valor máximo de AEP	17.5	17.2	13.5	14.2
Porcentaje de AEP inferior a 5%	55.9	57.4	72.1	72.1
Porcentaje de AEP superior a 10%	8.8	14.7	2.9	1.5

Fuente: Elaboración propia.

Cuadro 4
**VALOR DE R² Y SIGNIFICACIÓN DADA POR LA PRUEBA "T"
 PARA LOS TRES MODELOS DE REGRESIÓN MÚLTIPLE
 POR VARIABLE SINTOMÁTICA UTILIZADA, SEGÚN
 MÉTODO DE ESTIMACIÓN
 (NIVEL CANTONAL) ^a**

Variables sintomáticas incluidas	Correlación de razón		Correlación de tasa	
	R ²	Prueba "t"	R ²	Prueba "t"
Modelo 1	0.76		0.74	
Padrón electoral		0.00		0.00
Modelo 2	0.81		0.81	
Padrón electoral		0.00		0.00
Nacimientos		0.00		0.00
Modelo 3	0.82		0.82	
Padrón electoral		0.00		0.00
Nacimientos		0.00		0.00
Matrícula escolar		0.07		0.06

Fuente: Elaboración propia.

^a Véanse los modelos de regresión presentados en el cuadro 3.

3. Comparación con estimaciones obtenidas mediante el método de crecimiento vegetativo

Como se ha mencionado, existe una gran demanda de estimaciones poblacionales basadas en el crecimiento vegetativo de la población, específicamente de las producidas por la Dirección General de Estadística y Censos de Costa Rica. Por lo tanto, se decidió comparar las estimaciones de la DGEC en 1984 con los resultados obtenidos mediante la aplicación de las variables sintomáticas. En el cuadro 5 se presenta un resumen de los errores observados al comparar dichas estimaciones con la información censal correspondiente a los diferentes cantones.

Cuadro 5
**COMPARACIÓN DEL VALOR ABSOLUTO DEL ERROR PORCENTUAL
 (AEP) DE LAS ESTIMACIONES POBLACIONALES A NIVEL
 CANTONAL, CON CRECIMIENTO VEGETATIVO
 1984**

	Estimaciones bajo un supuesto de crecimiento natural	Estimaciones con variables sintomáticas ^a
Promedio de AEP	12.4	4.0
Valor máximo de AEP	60.7	12.5
Porcentaje de AEP inferior a 5%	28.4	72.1
Porcentaje de AEP superior a 10%	44.4	2.9

Fuente: Elaboración propia.

^a Los resultados presentados corresponden a la aplicación del método de correlación de tasa con el padrón electoral y los nacimientos.

Estos resultados evidencian grandes diferencias en términos de precisión; las más precisas son las estimaciones hechas sobre la base de variables sintomáticas. Si se analiza únicamente el promedio de *AEP*, se observa un incremento de más de ocho puntos porcentuales. Esto señala que, al menos en el caso de períodos bastante alejados del año censal, las estimaciones calculadas a partir de variables sintomáticas ofrecen una mejor alternativa que simplemente suponer un crecimiento vegetativo.

CONCLUSIONES

Los resultados del estudio señalan que son pocas las variables sintomáticas confiables de que se dispone. A pesar de ello, es posible realizar estimaciones poblacionales a nivel cantonal de una calidad aceptable.

El número de electores inscritos en el padrón electoral y los nacimientos son las variables que presentaron la más alta correlación (cercana al 0.90) con respecto a la población. Además, éstas fueron las

que ofrecieron las mejores estimaciones, especialmente con las técnicas en las que se utiliza regresión lineal. Si se usan estas variables como predictoras, sólo 3% de los cantones presenta una desviación del valor censal superior al 10%, con un error promedio cercano al 4%.

En el presente caso las estimaciones se efectuaron once años después del censo. Para períodos poscensales más cortos cabría esperar estimaciones más exactas, debido a que las proporciones y las tasas de las variables sintomáticas no presentarían cambios tan acentuados.

Las diferencias entre los cuatro métodos estudiados no son muy marcadas, por lo que se podría utilizar cualquiera de ellos sin que se produjeran mayores variaciones en cuanto a la calidad de las estimaciones; sin embargo, las técnicas de correlación de razón y correlación de tasa son evidentemente más precisas.

Un elemento importante que queda en evidencia en el estudio, y que merece ser analizado más en detalle, es la distribución de los errores en las estimaciones de acuerdo con el tamaño poblacional. Los modelos de estimación demuestran que, mientras más reducida es la magnitud de la población, menor es la precisión de las estimaciones. Este hecho se observa claramente en la disminución del promedio de *AEP* cuando se pondera por el tamaño de la población de cada área menor, así como en las diferencias, ya comentadas, entre provincias, cantones y distritos. Este fenómeno puede deberse a muchas causas. En primer lugar podría atribuirse a la calidad de la información sintomática, que se va deteriorando a niveles desagregados menores (Rincón, 1989). Otra causa podría ser el hecho de que los errores dependen de la calidad de los cómputos censales. Sin embargo, la información censal está sujeta a problemas de cobertura, que difieren en magnitud de un área menor a otra. Lógicamente, es de esperar que esas diferencias sean más marcadas en unidades político-administrativas más pequeñas y en poblaciones más reducidas (Wolter y Causey, 1991). Una tercera causa que podría explicar las diferencias en lo que respecta a los errores —y que constituyó un serio problema en el manejo de información—, es la variación de los límites geográficos producida por la creación constante de nuevas áreas. Esta inestabilidad obliga a llevar a cabo reagrupaciones o desagregaciones de las localidades pertinentes, lo que puede provocar errores en el manejo de los datos sobre población y de la información sintomática.

Una de las metas implícitas del estudio era el desarrollo de una metodología que permitiera realizar estimaciones poblacionales de mejor calidad que la obtenida bajo el supuesto de un crecimiento

vegetativo de la población. Al comparar las estimaciones resultantes de la aplicación de los cuatro métodos con las publicadas por la Dirección General de Estadística y Censos sobre 1984, se observaron grandes diferencias en términos de precisión. El promedio de *AEP* fue de 12.4%, porcentaje que supera en alrededor de 8.5 puntos porcentuales las estimaciones hechas en el estudio.

Pese a lo anterior, y debido a la falta de información censal, la prueba de precisión de las estimaciones poblacionales tuvo que realizarse considerando el año 1984. Esto explica la necesidad de repetir el estudio tan pronto como se realice un nuevo censo de población, a fin de observar la estabilidad de las variables sintomáticas y de los métodos de estimación.

El presente estudio permite concluir que la exactitud de un método no puede ser generalizada y que, aunque un determinado procedimiento de estimación sea adecuado, no asegura que su aplicación a diferentes situaciones conduzca a resultados que reflejen fielmente la realidad. Al analizar las estimaciones se deben considerar las hipótesis sobre 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.

BIBLIOGRAFÍA

- CELADE/DGEC/MIDEPLAN (Centro Latinoamericano de Demografía/ Dirección General de Estadística y Censos/Ministerio de Planificación Nacional y Política Económica) (1988), *Proyecciones Nacionales de Población, 1975-2025*, 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.
- Ericksen, Eugene P. (1974), "A regression method for estimating population changes of local areas", *Journal of the American Statistical Association*, vol. 69.
- (1973), "A method for combining sample survey data and symptomatic indicators to obtain population estimates for local areas", *Demography*, N° 10.
- Feeney, Donald, James Hibbs y Thomas Gillaspay (1995), "Ratio-correlation method", *Basic Methods for Preparing Small-Area Population Estimate*, Norfleet Rives y otros (comps.), Laboratorio de demografía aplicada, Departamento de Sociología Rural, Escuela de Agricultura y Ciencias Biológicas, Universidad de Wisconsin-Madison, Servicio de Extensión.
- Mandell, Marylou y Jeffrey Tayman (1982), "Measuring temporal stability in regression models of population estimation", *Demography*, vol. 19.
- Namboodiri, Krishnan (1972), "On the ratio-correlation and related method of subnational population estimation", *Demography*, vol. 9.
- O'Hare, W. (1980), "A note on the use of regression methods in population estimates", *Demography*, vol. 17.
- Rincón, Mesa, Manuel José (1989), *Sistema para elaborar proyecciones subnacionales, por sexo y grupos de edades, por el método de los componentes*, San José, Costa Rica, Centro Latinoamericano de Demografía (CELADE).
- Rives, Norfleet W. (1982), "Assessment of a survey approach", *Population Estimates: Methods for Small Area Analysis*, Everetts S. Lee y Harold F. Goldsmith (comps.), Beverly Hills, California, SAGE Publications.
- Serow, William y Norfleet Rives (1995), "Small area analysis: assessing the state of the art", *Basic Methods for Preparing Small-Area*

- Population Estimate*, Norfleet Rives y otros (comps.), Laboratorio de demografía aplicada, Departamento de Sociología Rural, Escuela de Agricultura y Ciencias Biológicas, Universidad de Wisconsin-Madison, Servicio de Extensión.
- Simpson, Stephen y otros (1996), "Updating small area population estimates in England and Wales", *Journal of the American Statistical Association*, vol. 159.
- Swanson, David (1980), "Improving accuracy in multiple regression estimates of population using principles from causal modelling", *Demography*, vol. 17.
- Swanson, David y Lucky Tedrow (1984), "Improving the measurement of temporal change in regression models used for County population estimates", *Demography*, vol. 21.
- Voss, Paul R. y otros (1995), "Censal ratio methods", *Basic Methods for Preparing Small-Area Population Estimate*, Norfleet Rives y otros (comps.), Laboratorio de demografía aplicada, Departamento de Sociología Rural, Escuela de Agricultura y Ciencias Biológicas, Universidad de Wisconsin-Madison, Servicio de Extensión.
- Wolter, Kirk y Beverley Causey (1991), "Evaluation of procedures for improving population estimates for small areas", *Journal of the American Statistical Association*, vol. 86.

MÉTODOS DE ESTIMACIÓN PARA ÁREAS PEQUEÑAS Y UNA APLICACIÓN A LA PREVALENCIA ANTICONCEPTIVA*

Alfredo Aliaga
Macro Internacional Inc.

RESUMEN

Durante la implementación del programa internacional de encuestas de demografía y salud (EDS) ha aumentado la demanda de información sobre áreas geográficas o grupos de población no contemplados en el diseño de esas encuestas. En el presente documento se pasa revista a los principales obstáculos técnicos que deben enfrentar las diversas metodologías de estimación para satisfacer esa demanda y se propone un proceso de estimación alternativo, derivado de la metodología de regresión.

* Documento presentado en la Conferencia conjunta IAOS-AFSA, Comisión Económica para África de las Naciones Unidas, Addis Abeba, Etiopía.

**ESTIMATION METHODS FOR SMALL AREAS
AND THEIR APPLICATION IN MEASURING THE USE
OF CONTRACEPTIVES**

ABSTRACT

During the implementation of the Demographic and Health Survey (DHS) Programme, demand has been growing for information on geographic areas and population groups that have not been defined in the design of these surveys. This article reviews the main technical obstacles that the various estimation methodologies will have to overcome in order to meet that demand and proposes an alternative regression-based estimation methodology.

INTRODUCCIÓN

Aunque se pueden obtener *estimaciones directas* de las características sociodemográficas a nivel nacional en todas las encuestas por muestreo de las Encuestas Demográficas y de Salud (EDS), generalmente no es posible lograr *estimaciones directas para áreas pequeñas*, ya que el tamaño de las muestras correspondientes a esas áreas generalmente no es suficiente para que aquellas sean representativas —es decir, el tamaño adoptado para la muestra nacional es insuficiente para realizar estimaciones para los dominios pequeños. Del mismo modo, el nivel de clasificación utilizado para el diseño de las muestras nacionales no es adecuado para las áreas pequeñas. Sin embargo, en los últimos tiempos, los directores de programas y los encargados de la formulación de políticas de los países en desarrollo han manifestado interés en la estimación de indicadores sociodemográficos básicos para las áreas pequeñas a fin de definir objetivos, asignar recursos y supervisar el funcionamiento de los programas de atención de salud y de planificación familiar.

La realización de encuestas nacionales que también sean representativas de las áreas geográficas de nivel inferior es posible, pero no puede ser viable desde el punto de vista de los costos. Además, los errores no vinculados con el muestreo suelen aumentar a medida que crece el tamaño de la muestra. Una solución alternativa es utilizar las *estimaciones indirectas* para las áreas pequeñas. En el presente estudio se pasa revista a los diversos enfoques utilizados en las estimaciones de las áreas pequeñas y de los principales problemas que presentan.

1. FUENTES DE DATOS

Censos de población. Los censos de población —realizados generalmente cada 10 años— proporcionan los valores de base para obtener algunos indicadores sociodemográficos. En los países en que se realizan las EDS, se puede utilizar la información censal para determinar las características básicas de las áreas pequeñas y luego aplicar diversos procedimientos para elaborar las estimaciones estadísticas correspondientes a las áreas pequeñas seleccionadas. La información censal también puede utilizarse para determinar los factores de ponderación censal para las características no recopiladas en el censo y como antecedente en los procedimientos de modelización.

Registros administrativos. Los registros administrativos, como las estadísticas vitales o los registros de salud, son una fuente importante de información para las estadísticas de áreas pequeñas. Sin embargo, el programa de las EDS no los ha utilizado ampliamente porque en muchos casos no es fácil acceder a esos registros en los países en que se realizan las encuestas. Además, los registros administrativos se realizan más bien con fines programáticos que estadísticos, de modo que las definiciones individuales de las fuentes generalmente no son compatibles y no pueden combinarse con facilidad con los datos que poseen definiciones estandarizadas, como los datos de las EDS. Así, el uso de registros administrativos como fuente de información para las estadísticas de áreas pequeñas puede ser limitado.

Grandes encuestas de hogares. A menudo, las grandes encuestas de hogares están a cargo de las oficinas de estadística nacionales de los países en que se realizan las EDS. Cuando se dispone de datos correspondientes a las primeras, éstos pueden proporcionar información de base sobre las estadísticas de áreas pequeñas.

Las EDS. El proyecto de Encuestas de Demografía y Salud (EDS) proporciona información sobre algunos indicadores sociodemográficos y de salud para la población total, para las poblaciones urbanas y rurales y para algunas poblaciones regionales importantes. Así, las EDS, en forma individual o combinadas con otras bases de datos, permiten obtener los datos sociodemográficos y de salud necesarios para las estimaciones de áreas pequeñas.

2. PROBLEMAS FUNDAMENTALES DE LAS ESTIMACIONES DE ÁREAS PEQUEÑAS

Cualquier encuesta por muestreo se diseña para obtener estimaciones sobre dominios de gran tamaño. Aunque los problemas relacionados con las áreas pequeñas son críticos, resultaron menos polémicos para los dominios principales. Puede encontrarse un buen análisis sobre este tema en *Small Areas Statistics: An International Symposium* (Platek y otros, 1987).

Definiciones y conceptos. A veces no existe una definición común o un mismo conjunto de definiciones y conceptos en las diversas fuentes de datos. Por ejemplo, puede haber diferencias en las condiciones que deben reunir los encuestados para cada fuente de datos, en los procedimientos utilizados para la recopilación de datos y en la manera en que el método de estimación se aplica a la base de datos. Tal disparidad hace difícil combinar las fuentes de datos. Estos problemas son más graves en los países en desarrollo.

Cambios en las políticas de los gobiernos. Los cambios en las políticas de los gobiernos pueden incidir en la producción de datos de las fuentes. Estos cambios se verán reflejados en la población que abarca la fuente, el grado de cobertura de diversos temas y la calidad de los datos obtenidos, asociada al tiempo y los recursos asignados a la recopilación de datos.

Problemas de los procedimientos de estimación. Los procedimientos de estimación están supeditados a la situación relativa a los factores mencionados. El tipo de estimación dependerá de las condiciones de la información del país de que se trate. En muchos países en desarrollo es recomendable recurrir a procedimientos sencillos en lugar de utilizar métodos complejos. Para estos últimos se requiere personal muy capacitado, una mayor asignación de recursos y un procesamiento de datos intensivo.

Privacidad. Como existen diferencias entre las fuentes y es preciso combinar información de varias bases de datos, a menudo hay que analizar registros individuales o grupos de registros. El problema de la privacidad y sus consecuencias es una preocupación legítima del público en general y los investigadores deben establecer procedimientos claros para asegurar la privacidad de la información.

Tamaño. El censo de población es la mejor fuente para la estimación de una estadística básica referida a un dominio, cualquiera sea la magnitud de éste. Sin embargo, para estadísticas más complejas,

recopiladas en estudios limitados —y no en censos— no pueden realizarse estimaciones de estas estadísticas para todos los casos en razón de las limitaciones del número de casos, ya que puede existir carencia de datos o muy pocos datos. Las encuestas por muestreo pueden desempeñar un papel importante en la complementación de esa información. Sin embargo, no permiten obtener estimaciones para dominios de cualquier tamaño. Generalmente, la muestra para una encuesta se diseña con el fin de calcular estimaciones para dominios de gran magnitud. La definición del área pequeña está relacionada con los objetivos del diseño de la muestra y con su tamaño.

En general, el programa EDS proporciona estimaciones de los parámetros demográficos básicos correspondientes al país en su conjunto, a las áreas urbanas y rurales como subgrupos distintos y a las principales regiones del país. Las áreas de menor tamaño que las regiones principales apenas están representadas en las EDS y el tamaño de la muestra termina siendo nulo o muy pequeño. En el programa de EDS, el término *área pequeña* se limita al nivel geográfico inmediato inferior —generalmente una provincia o un distrito— cuyos datos no se publican. Actualmente, el programa de EDS no prevé realizar estimaciones para las áreas pequeñas.

Estimación. La estimación del dominio forma parte de todas las encuestas por muestreo. El tipo de procedimiento de estimación depende del diseño de la muestra. Como es difícil medir el sesgo, los procedimientos que permiten obtener estimaciones no sesgadas son más aceptables que los procedimientos que dan por resultado estimaciones sesgadas. Algunos autores prefieren los procedimientos de estimación consistentes porque cuando el tamaño de la muestra llega a ser suficientemente grande, el valor de la estimación es más confiable para el dominio, es decir, el valor del sesgo tiende a ser pequeño. Algunos autores tratan de utilizar esa consistencia para asegurar que las propiedades de la muestra sean convenientes, al menos cuando ésta es grande.

Otro método de clasificación general de los procedimientos de estimación se basa en el nivel de probabilidad muestral del diseño de la muestra o en el uso de datos auxiliares de otras fuentes. Se puede hacer una *estimación directa* usando los datos de la encuesta, pero sólo las unidades correspondientes al dominio del área pequeña. Otro tipo de estimación es una modificación de la estimación directa, denominada *estimación directa modificada*. En ella se utilizan datos de la encuesta principal —de otros dominios, tanto para la variable de estudio como

para la auxiliar— pero se conservan las propiedades del diseño, como la consistencia y la ausencia de sesgo. Finalmente, existe un número creciente de investigaciones sobre las *estimaciones indirectas (o modelización)* que utilizan datos externos —tanto para las variables de estudio como las auxiliares— que no pertenecen al dominio o al período correspondientes al estudio, con o sin consideración de las propiedades de diseño de la encuesta.

Evaluación. Una forma ampliamente aceptada de evaluar cualquier estimación es determinar su error relativo. Este valor se calcula como el cociente entre el error estándar de la estimación y el valor de ésta. Un problema fundamental para cualquier encuesta por muestreo es decidir si es más conveniente evaluar la estimación de un gran número de variables usando estimadores de diseño o la de un número reducido de estimaciones indirectas realizadas porque el tamaño de la muestra era pequeño.

Los procedimientos que permiten obtener estimaciones consistentes presentan otros problemas. Las características de estas estimaciones, calculadas a partir de una muestra de gran tamaño, son adecuadas, pero no ocurre lo mismo cuando el tamaño de la muestra es pequeño. Por otra parte, una de las dificultades más importantes radica en obtener medidas confiables de calidad —errores de muestreo o errores cuadráticos medios— cuando se usan estos procedimientos de estimación para una encuesta determinada.

Un problema fundamental en ciertos procedimientos de modelización para el cálculo de las estimaciones es que el modelo no sea válido para algunas áreas pequeñas. Puede ocurrir algo semejante con las estimaciones directas: si bien la muestra puede tener un tamaño razonable, es posible que el valor de la estimación no sea confiable debido a las variaciones muestrales.

Otra dificultad en la estimación de áreas pequeñas está vinculada con el uso de información correspondiente a distintos períodos. Si el investigador quiere analizar los cambios en el tiempo de una determinada variable habrá ciertas dudas sobre la confiabilidad de esa estimación. Este problema es menos evidente cuando sólo se requiere una estimación puntual en un momento determinado.

Idealmente, el objetivo es desarrollar un procedimiento razonable de estimación que sea apropiado a todas las estimaciones posibles del dominio según un procedimiento de calidad determinado. Al respecto, se sugiere realizar un estudio de validación para medir la eficacia de cada valor directo en comparación con un valor indirecto bajo varias

condiciones. Sin embargo, la realización de un estudio de esta naturaleza exige un esfuerzo muy grande.

Diseño muestral. El diseño muestral satisface las necesidades de los dominios principales. Sin embargo, se debe considerar la posibilidad de contar con enfoques apropiados para otros dominios en los que la estimación se considera posible en un futuro cercano. Los procedimientos de estimación para los dominios planificados están bien establecidos, pero resultan más difíciles para los dominios no planificados. Para resolver este problema es preciso encontrar fórmulas de transacción en el diseño de la muestra. Las soluciones parciales comprenden modificaciones en las estrategias de afijación del tamaño muestral, del nivel de agrupación y de la estratificación de la muestra.

Si se conoce el área pequeña de antemano, la afijación de la muestra se puede modificar mediante el diseño de la muestra —y tratar el área como dominio planificado— aumentando el tamaño de ésta. (Esto supone disponer de suficientes recursos humanos y financieros.) Si mediante el estudio sólo se desean obtener estimaciones nacionales, la afijación proporcional de la muestra puede ser suficiente. Sin embargo, si la muestra se utilizara para obtener estimaciones provinciales o de distrito, entonces se podrían obtener tamaños de muestra comparables entre provincias fijando un tamaño de muestra igual para cada una a fin de lograr el mismo nivel de confiabilidad. Si se aumenta el tamaño de las muestras de las provincias según las necesidades, se obtendrá un tamaño muestral mayor al que correspondería al país en su conjunto mediante una afijación proporcional. Podría producirse una situación similar si primero se examina una estimación provincial y luego una estimación por distritos de una misma provincia. En este caso, el diseño de la muestra tendrá que conciliar la confiabilidad de los diversos niveles del dominio. En general, será necesario encontrar una solución de compromiso entre distintos tipos de afijación: proporcional o igualitaria, óptima o proporcional, óptima o igualitaria, etc.

Otro factor que puede utilizarse para mejorar las estimaciones de áreas pequeñas es el nivel de clasificación. En general, el nivel apropiado para las estimaciones de los dominios nacionales y subnacionales no es adecuado para la representación de las áreas pequeñas. Teniendo en cuenta la magnitud de los recursos disponibles para la encuesta, una solución de compromiso es disminuir el nivel de clasificación en las primeras etapas del diseño de la muestra. En el programa EDS, por lo general, los niveles de clasificación utilizados

son adecuados para los niveles nacionales, subnacionales y regionales. Sin embargo, para mejorar el tamaño muestral para las áreas pequeñas, debe incorporarse al diseño de la muestra un número mayor de unidades más pequeñas en la primera etapa.

La estratificación de las muestras puede utilizarse para determinar áreas pequeñas más representativas disminuyendo el tamaño de los estratos. Si al mismo tiempo se disminuye el nivel de clasificación, se mejorarán las estimaciones para los dominios no planificados.

Redefinición de las áreas pequeñas. Si no se logra mejorar las estimaciones del dominio pequeño aumentando el nivel de la afijación o disminuyendo el nivel de clasificación, una solución intermedia sería redefinir el dominio colapsando áreas pequeñas cercanas entre sí y realizando una afijación de tamaño muestral suficiente en ese nivel. De este modo se mejoraría el proceso de estimación confiando en que el área pequeña redefinida permitirá obtener información similar a la del área original. Este enfoque, conveniente en la práctica, puede ser objetable, pues supondría que la condición de homogeneidad del dominio redefinido también sería válida para el dominio pequeño original.

Si el país se propone modificar los límites de los dominios administrativos, de modo que también se vean afectados los límites de las áreas pequeñas en un futuro cercano, podría plantearse un problema de redefinición. Esa posibilidad puede tenerse en cuenta en el diseño, redefiniendo áreas unitarias estables —no afectadas por el cambio— y seleccionables, que más adelante se podrían agrupar para redefinir el dominio pequeño.

Aunque los problemas que acabamos de enumerar no son exhaustivos o exclusivos, pueden interactuar entre sí impidiendo, de hecho, una solución al problema de la redefinición de las áreas pequeñas —especialmente si el tamaño de la muestra no se diseñó inicialmente para realizar estimaciones de dominios pequeños. Si en el proceso de diseño no se tuvo en cuenta la necesidad de realizar una estimación de área pequeña, en una etapa posterior del análisis puede ser necesario desarrollar procedimientos de estimación para los dominios pequeños.

3. PROCEDIMIENTOS DE ESTIMACIÓN

Lo que sigue es una presentación de algunos enfoques utilizados para la estimación de dominios pequeños. No es intención del autor pasar revista exhaustivamente a los procedimientos de estimación de áreas pequeñas. Una buena descripción de tales procedimientos puede encontrarse en *Small area statistics: An international symposium* (Platek y otros, 1987).

Estimación directa. Como ya se ha analizado, la estimación directa de áreas pequeñas se basa en los datos de encuestas provenientes solamente del dominio del área pequeña. Un procedimiento aceptable en la mayoría de los casos se basa en el diseño, especialmente si aquel es prácticamente insesgado. Sin embargo, ese tipo de diseño no es satisfactorio cuando el tamaño de la muestra es pequeño, es decir, cuando los errores de muestreo de las estimaciones son grandes. Es preciso destacar enfáticamente que la estimación directa *sólo debe ser utilizada* cuando el tamaño de muestra es suficientemente grande como para permitir una estimación confiable.

La fórmula general para la estimación del cociente de una estimación directa viene dada por:

$$r = \frac{\sum_i w_i y_i}{\sum_i w_i x_i} \quad [1]$$

donde

w_i es el factor de ponderación de diseño para la unidad i ;
 y_i es el valor de y para la unidad i ; y
 x_i es el valor de x para la unidad i .

Si X es una variable de enumeración, el estimador del cociente se reduce a la media ponderada de Y , y puede expresarse como:

$$r = \frac{\sum_i w_i y_i}{\sum_i w_i} \quad [2]$$

Cuando se conoce el tamaño N de la población, la estimación directa de una expansión total viene dada por:

$$\hat{Y} = N \frac{\sum_i w_i y_i}{\sum_i w_i} \quad [3]$$

En general, si se conoce el valor total de la variable X para todos los estratos, podemos definir el valor total ampliado como:

$$\hat{Y} = \sum_h X_h r_h \quad [4]$$

donde

X_h es el valor total conocido de la variable auxiliar en los estratos h .

Puede obtenerse una estimación directa de la regresión como estimador aproximado de diseño no sesgado. Viene dada por:

$$Y_{reg.a}^* = Y_a^* + \beta_a^* (X_a - X_a^*) \quad [5]$$

donde:

$Y_{reg.a}^*$ es el valor total expandido predicho en el dominio pequeño a ;

β_a es el coeficiente de la regresión en el dominio pequeño a ;

donde β_a se calcula como:

$$\beta_a^* = \frac{\sum_{i \in s_a} v_i^{-1} \omega_i Y_i X_i}{\sum_{i \in s_a} v_i^{-1} \omega_i X_i X_i} \quad [6]$$

donde el valor del subíndice i corresponde a la i -ésima unidad en el dominio de la muestra a ;

X_a es el valor total X para el dominio pequeño a ;

Y_a^* es el valor de expansión de Y en el dominio pequeño a ; y

X_a^* es el valor de expansión de X en el dominio pequeño a .

Se puede obtener un estimador directo aproximado modificado sustituyendo el valor del coeficiente para un valor calculado para toda la muestra, que podemos llamar $Y_{\text{reg-mo},i}^*$

Estimación indirecta. Este tipo de estimación está relacionado con el procedimiento de estimación sintético, que se basa en la hipótesis de que el área pequeña es similar en algún sentido a otra área, a menudo un área más grande que la contiene. Aunque este tipo de estimación tendría una varianza más pequeña, puede resultar sesgada si no se cumple la hipótesis.

$$Y_i = \sum_j \left(\frac{N_{ij}}{N_i} \times y_j \right) \quad [7]$$

donde

Y_i = el valor total estimado o el valor del cociente para la i -ésima área pequeña;

y_j = el valor total observado o el valor del cociente para la j -ésima categoría auxiliar;

N_{ij} = el número estimado de unidades en la j -ésima categoría auxiliar, para la i -ésima área, tal como se observa en la fuente externa;

N_i = el número estimado de unidades en la i -ésima área, tal como se observa en la fuente externa, y

N_{ij}/N_i = las ponderaciones de ajuste.

Entre los métodos indirectos para obtener estimaciones de área pequeña cabe mencionar los que se basan en el uso de un modelo y, en especial, los que utilizan una regresión. Generalmente, la ventaja de este último enfoque es que logra cierta estabilidad en el modelo usando información cercana o circundante al dominio pequeño.

Aliaga y Muhuri (1994) han desarrollado un procedimiento de modelización (con varias características deseables) que permite estimar la tasa de prevalencia del uso de anticonceptivos como variable dependiente relacionada con once variables independientes:

- Cada variable es categórica a nivel individual (dos o más categorías).
- Cada variable fue transformada en una variable binaria; la dicotomización se basó en la *maximización de la capacidad de discriminación sobre el uso de anticonceptivos*. Se creó una variable continua (proporciones o porcentajes) para la unidad del nivel inmediato superior (conglomerado) para cada variable, *es decir, cada una tenía la misma escala y era continua*. Alrededor de cada dominio del área pequeña, se crearon grupos con áreas vecinas, en forma circular, de modo que cada grupo tuviera alrededor de 30 conglomerados. Cada grupo comprende de dos a cinco dominios pequeños, *es decir, que si se circunda el dominio pequeño en forma circular se obtendrá información sobre todas las áreas vecinas que puedan tener un comportamiento similar*.
- Se crearon modelos de regresión a nivel de conglomerado para cada grupo que tuviera *estabilidad* (cada grupo tenía alrededor de 30 conglomerados).
- La distribución del cociente entre la media muestral y el error de muestreo estimado se puede aproximar mediante la distribución normal.
- Se estimó la tasa de prevalencia para cada grupo que circundaba el área de dominio pequeño usando los valores pronosticados correspondientes, a nivel de conglomerado.
- La estimación final para el dominio pequeño se obtenía promediando las estimaciones de todos los grupos.

Cualquiera sea el modelo utilizado, pueden plantearse varios problemas o restricciones. Por ejemplo, la estimación se basa en las hipótesis del modelo, que pueden no cumplirse. Asimismo, la estimación se debe restringir a determinados tipos de variables.

Estimadores combinados. Cuando se han calculado al menos dos tipos de estimaciones distintas, es posible también calcular una estimación que sea una combinación lineal de ambas. Por ejemplo, si hay una estimación directa y otra indirecta, la nueva estimación se puede calcular como sigue:

$$r_c = \alpha r_d + (1 - \alpha) r_i \quad [7]$$

donde

α tiene un valor entre 0 y 1;
 r_c es el estimador combinado;
 r_d es el estimador directo; y
 r_i es el estimador indirecto.

El valor de a puede considerarse un factor de ponderación asignado al estimador directo. Por lo tanto, el término $(1 - \alpha)$ es la ponderación para la estimación indirecta. Entre las distintas soluciones posibles, una de ellas se obtiene cuando a es proporcional a la precisión de la estimación directa (la inversa del error de muestreo), es decir, $(1 - \alpha)$ es proporcional a la precisión de la estimación indirecta. La estimación combinada siempre será relativamente más cercana a la que tiene más precisión y tendrá un valor intermedio entre ambas.

Otros procedimientos. Hay varios otros procedimientos de estimación basados en principios estadísticos, como el enfoque bayesiano, cuya forma es similar al estimador combinado, siendo a un valor probabilístico. Si se dispone de información *a priori* para estimar una distribución de ese tipo (a), puede calcularse una estimación razonable para el dominio pequeño. Aunque este principio es bastante interesante, una de las dificultades principales es la necesidad de establecer una distribución *a priori*. El uso del enfoque bayesiano mediante el muestreo de Gibbs (generando una secuencia de Gibbs de variables aleatorias de distribuciones condicionales, que exige un uso intensivo de sistemas computarizados) puede servir para superar este problema empírico generando la distribución *a posteriori* de la estimación. El muestreo de Gibbs plantea problemas importantes en cuanto a la puesta en práctica y la comparación de las diversas maneras de extraer información de la secuencia de Gibbs.

4. APLICACIONES

Cualquier procedimiento de estimación que se base en el diseño de la muestra es más aceptable que un procedimiento que no tenga en cuenta el diseño. De éstos, cualquier procedimiento que se base en el diseño y sea insesgado (o aproximadamente insesgado) es más aceptable que un procedimiento sesgado. Para la estimación directa se debe utilizar, siempre que sea posible, un procedimiento basado en el diseño e insesgado (o aproximadamente insesgado). Sin embargo, a medida que el

nivel del dominio se hace más pequeño, este tipo de enfoque pierde fiabilidad. Aunque existen restricciones para los dominios pequeños, el procedimiento de estimación directa da una idea general de la estimación para esos dominios críticos (un tamaño de muestra pequeño, es decir, un número reducido de conglomerados).

En las EDS, para cualquier dominio que tenga 30 o más conglomerados, el procedimiento de estimación directa permitirá obtener una estimación homogénea y fiable y la probabilidad de que sea inadecuada será baja (5%). Para los dominios que tienen entre 10 y 30 conglomerados, el procedimiento de estimación directa puede ser útil si se combina con otros resultados. Para los dominios de 10 conglomerados o menos, el procedimiento de estimación directa es muy poco confiable. Tales estimaciones no deben utilizarse directamente, sino que se pueden combinar con otros tipos de estimación.

En el caso de la estimación provincial de la República Dominicana (véase el cuadro 1), y la estimación distrital de Kenia (véase el cuadro 2), se determinó que:

1. En las estimaciones del Distrito Nacional y la provincia de Santiago en la República Dominicana y del distrito de Nairobi en Kenia, usando alrededor de 30 conglomerados o más, se obtuvieron valores absolutamente coherentes utilizando distintos enfoques. Por lo tanto, para la prevalencia del uso de anticonceptivos la estimación directa debe usarse sin más.

2. Para cada una de las provincias de la República Dominicana y cada uno de los distritos de Kenia, en que las muestras tienen entre 10 y 20 conglomerados, el enfoque de la regresión arroja resultados absolutamente coherentes con las estimaciones combinadas. Por lo tanto, cuando la diferencia absoluta entre la regresión y las estimaciones combinadas es menor que el 10% de la regresión, la alternativa razonable recomendada es la estimación por regresión. Quizás deba usarse la estimación combinada cuando la diferencia supere el 10%.

3. Para cada una de las provincias de la República Dominicana y cada uno de los distritos de Kenia, con una muestra de 10 conglomerados o menos, el valor medio entre la regresión y las estimaciones sintéticas permitiría obtener una estimación más razonable.

Cuadro 1
**ESTIMACIONES DIRECTA, INDIRECTA, POR REGRESIÓN Y COMBINADA
 DE LA PREVALENCIA DEL USO DE ANTICONCEPTIVOS POR LAS
 MUJERES CASADAS: REPÚBLICA DOMINICANA, 1991**

Región y provincia	Estimación directa ^a			Estimación por regresión		
	Valor	Error muestral	Estimación indirecta ^b	Valor	Error muestral	Estimación combinada
Región 0	60.7					
Distrito Nacional	60.7	2.0	60.6	61.7	1.8	61.2
Región I	50.8					
Peravia	48.7	5.8	51.5	54.6	3.0	52.6
San Cristóbal	55.9	3.7	53.7	52.3	2.6	53.8
Monte Plata	43.1	3.9	51.4	52.3	3.5	47.9
Región II	61.0					
Santiago	62.1	3.5	62.1	60.7	4.4	61.4
Puerto Plata	62.8	5.3	62.0	58.6	2.1	59.8
La Vega	59.4	4.4	61.5	58.0	2.0	58.4
Españillat	56.2	9.1	62.0	57.3	3.5	57.0
Monseñor Nouel	61.7	1.5	61.2	56.7	2.0	59.6
Región III	57.4					
Salcedo	53.4	5.6	57.1	56.1	3.4	55.1
Duarte	57.2	4.5	56.1	55.3	2.3	55.9
María T. Sánchez	65.9	7.5	57.8	57.9	1.6	59.3
Samaná	44.6	10.2	55.0	57.9	1.6	56.1
Sánchez Ramírez	56.7	8.2	58.5	53.7	4.3	54.7
Región IV	47.1					
Barahona	51.9	2.7	47.9	50.0	3.3	51.0
Pedernales	48.3	9.6	45.1	50.0	2.4	49.7
Bahoruco	38.8	6.5	47.7	43.4	4.8	41.4
Independencia	44.8	9.1	46.2	49.1	2.5	48.2
Región V	50.6					
La Romana	44.0	7.9	50.1	50.1	4.0	48.0
La Altagracia	57.5	3.8	51.4	49.8	4.1	53.8
El Seybo	46.5	9.7	51.0	49.5	3.4	48.7
San P. de Marcorís	54.2	3.5	52.3	50.8	3.3	52.5
Hato Mayor	58.0	7.0	51.9	49.6	2.7	51.9
Región VI	39.7					
San Juan	34.1	8.2	39.5	41.5	6.6	38.2
Azua	52.3	4.5	39.0	52.7	3.2	52.5
Elías Piña	32.1	5.9	39.9	40.5	8.0	35.7
Región VII	58.6					
Valverde	53.9	2.4	59.1	58.4	2.9	55.9
Santiago Rodríguez	73.6	7.4	62.7	55.1	4.9	62.5
Dajabón	51.8	5.6	56.5	55.8	6.3	53.7
Montecristi	62.8	4.4	56.3	59.3	2.6	60.6

^a Datos de la EDS de 1991.

^b Datos de la EDS de 1991 y datos de la Encuesta de Hogares Ampliada de 1991.

Cuadro 2
**ESTIMACIONES DIRECTA, INDIRECTA, POR REGRESIÓN Y COMBINADA
 DE LA PREVALENCIA DEL USO DE ANTICONCEPTIVOS POR LAS
 MUJERES CASADAS: KENIA, 1992**

Región y provincia	Estimación directa ^a			Estimación por regresión		
	Valor	Error muestral	Estimación indirecta ^b	Valor	Error muestral	Estimación combinada
Nairobi						
Nairobi	33.5	2.1	34.9	38.5	3.6	35.3
Central	39.5					
Kiambu	37.3		38.1	37.9	3.2	
Kirinyaga	54.2	3.2	38.5	44.4	2.0	48.2
Muranga	32.1	5.4	39.2	38.8	2.5	36.7
Nyandarua	39.2		38.4	40.4	3.2	
Nyeri	40.8	3.2	39.0	43.3	2.8	42.1
Costera	18.1					
Kilifi	10.8	1.7	17.2	15.2	2.6	12.5
Kwale	16.2		17.0	17.3	4.3	
Mombasa	24.5	3.6	17.0	15.0	2.0	18.4
Taita	30.7		18.1	29.0	1.9	
Oriental	40.2					
Embu	47.2		40.0	42.1	2.4	
Kitui	41.3		39.6	41.6	2.7	
Machakos	40.4	2.9	40.0	41.0	1.7	40.8
Meru	36.3	3.5	39.9	43.4	3.8	39.7
Nyanza	13.8					
Kisii	21.5	2.6	13.5	16.9	1.4	18.5
Kisumu	17.8	2.8	13.3	14.8	1.3	15.8
Siaya	8.5	2.2	13.7	10.1	1.5	9.5
Nyanza meridional	6.1	1.4	13.2	11.2	1.0	9.1
Valle del Rift	29.6					
Baringo	12.5		29.4	27.1	3.1	
Elgeyo Marak	16.7		29.4	15.0	2.8	
Kajiado	52.0		28.8	41.5	4.6	
Kericho	24.1	2.6	29.1	19.3	1.6	21.1
Laikipia	68.5		29.4	45.0	4.5	
Nakuru	47.2	5.4	29.2	34.0	5.2	40.5
Nandi	16.7		29.3	18.0	2.2	
Narok	22.6		28.8	23.3	4.5	
Trans Nzoia	28.9		29.2	15.1	2.5	
Uasin Gishu	13.4	2.5	29.2	15.7	2.3	14.6
Pokot Occidental	0.0		29.1	n.d.	n.d.	
Occidental	13.7					
Bungoma	9.4	1.5	13.2	11.8	1.8	10.5
Busia	16.1		13.2	11.7	2.2	
Kakamega	14.9	2.0	n.d.	14.4	1.7	14.6

^a Datos de la EDS de 1989.

^b Datos de la EDS de 1989 y censo de población de 1990.

Nota: Los errores de muestreo de las estimaciones directas sólo se calcularon para los distritos de más de 9 unidades primarias de muestreo (UPM).

n.d.: no se dispone de datos.

5. CONCLUSIONES

Diversos estudios indican que no existe un enfoque único para obtener estimaciones para los dominios pequeños cuyos resultados sean satisfactorios en todos los casos. Las propiedades de una estimación determinada dependerán del tamaño y la homogeneidad del dominio de área pequeña, de la disponibilidad de información externa, del carácter continuo o discreto de la variable de interés, de su idoneidad para explicar cualquier relación hipotética y del nivel de capacidad para discriminar las ocurrencias de un evento. Un enfoque adecuado para una situación determinada puede no serlo para otra.

Siempre puede considerarse la posibilidad de utilizar otro tipo de estimaciones (en forma individual o combinada). Para los dominios que tienen un número crítico de conglomerados, la estimación definitiva sólo debe darse después de realizar una evaluación del tipo de variable y de las condiciones de entorno del dominio pequeño.

Los procedimientos combinados de estimación deben utilizarse después de realizar una evaluación cuidadosa y no en forma automática, es decir, usando criterios puramente matemáticos. La información disponible en el sector público puede haberse utilizado con fines muy diversos. Por lo tanto, la combinación de distintas fuentes de información debe hacerse con prudencia para evitar controversias.

BIBLIOGRAFÍA

- Aliaga, Alfredo y Thanh Le (1991), "Methodology for small area estimation with DHS samples", *Institute for Resource Development. Demographic and Health Surveys World Conference; proceedings*, vol. 1, Columbia, Maryland, IRD/Macro International Inc.
- Aliaga, Alfredo y Pradip K. Muhuri (1994), "Methods of estimating contraceptive prevalence rates for small areas: applications in the Dominican Republic and Kenya", DHS Methodological Reports, N° 3, Calverton, Maryland, Macro International Inc.
- González, M.E. (1973), "Use and evaluation of synthetic estimates", *Proceedings of the American Statistical Association (Social Statistics Section)*, Washington, D.C., Asociación Americana de Estadística.
- Platek, R. y otros (comps.) (1987), *Small area statistics: An international symposium*, Nueva York, Wiley & Sons.

ESTIMACIÓN DE LA DEMANDA DE PLANIFICACIÓN FAMILIAR PARA LOS DEPARTAMENTOS Y PROVINCIAS DE PERÚ

Guillermo Vallenás Ochoa
Universidad Nacional Mayor de San Marcos

RESUMEN

Este trabajo busca generar tasas de prevalencia anticonceptiva para divisiones menores (provincias) a la vez que cuantificar los subgrupos de mujeres consideradas como usuarias de métodos de planificación familiar, para lo que se aplicará una metodología que combina datos de la ENDES 1992-1993 con datos del Censo de 1993. El procedimiento consiste en seleccionar un conjunto de variables que se relacionen al uso de anticonceptivos, que hayan sido investigadas tanto en la ENDES como en el censo. Mediante la aplicación de técnicas de análisis multivariado a los datos de la encuesta, se determinan las dos variables más importantes en relación a la práctica anticonceptiva. Se realizan estimaciones regionales de la demanda de planificación familiar sobre la base de la ENDES para estas variables y sus categorías. Con los datos censales se generan tablas de contingencia de las mujeres unidas de acuerdo a las variables seleccionadas y sus respectivas categorías, a nivel de provincia, para usar las como ponderadores de la estimación muestral. La estimación de la tasa de prevalencia provincial resulta de combinar linealmente las tasas de prevalencia regionales con los ponderadores provinciales. Así se obtienen resultados por provincias para las 13 regiones administrativas que considera la ENDES II. Las estimaciones provinciales están expresadas en tasas o proporciones de mujeres unidas para cada grupo componente de la demanda de métodos anticonceptivos.

ESTIMATING DEMAND FOR FAMILY PLANNING SERVICES IN THE DEPARTMENTS AND PROVINCES OF PERU

ABSTRACT

The purpose of this study is to generate contraception prevalence rates for small administrative units (provinces) and to quantify subgroups of the female population who are considered to be users of family planning methods. The methodology employed combines regional and national data from Demographic Health Surveys (ENDES) 1992-1993 and data from the 1993 Population and Housing Census. The procedure first selects a set of variables related to the use of contraceptives, which have been researched by both the ENDES and the census. Then multivariate techniques are applied to the survey data in order to identify the two most important variables of contraceptive use. Regional estimates of demand for family planning are drawn on the basis of the ENDES results for these variables and their respective categories. The census data are used to generate contingency tables showing the number of women who are in formal and consensual unions, classified according to the variables identified and their respective categories at the provincial level. The provincial prevalence rate is then estimated by means of a linear combination of regional prevalence rates with provincial weightings obtained from the contingency tables. Thus, results are obtained by province, for the 13 administrative regions covered by ENDES II. Provincial estimates are expressed as rates or proportions of women in formal and consensual unions for each component group of demand for contraceptive methods: unsatisfied demand, actual use and failure of the method, the sum of which represents total demand.

INTRODUCCIÓN

Uno de los objetivos de la Ley de Política Nacional de Población de Perú es “promover y asegurar la decisión libre, informada y responsable de las personas y las parejas sobre el número y espaciamiento de los nacimientos, proporcionando para ello los servicios educativos y de salud para contribuir a la estabilidad y solidaridad familiar y mejorar la calidad de vida”. Para esto, el Estado impulsa la realización de programas de planificación familiar que involucren actividades de educación, información y prestación de servicios en acción coordinada entre las instituciones públicas y no públicas. La eficiencia de estas acciones en una población cuya principal característica es su composición sociodemográfica heterogénea y con grandes desequilibrios regionales, depende del nivel de cobertura territorial, así como de la calidad de la planificación y de la prestación de los servicios. Esto significa llegar con los servicios y recursos necesarios a niveles departamentales, provinciales y distritales.

Un problema fundamental en la programación de este tipo de actividades y la estimación de los recursos de planificación familiar en unidades geográficas desagregadas como las mencionadas, es la carencia de información. En los registros administrativos del Ministerio de Salud (MINSA) se toma en cuenta únicamente la cobertura de sus propios programas, dejando de lado las realizadas por el sector privado. Ante esta limitante, también se podrá recurrir, como fuente de información, a las encuestas de hogares, entre otras la Encuesta Demográfica y de Salud Familiar realizada en 1991-1992 (ENDES II). Sin embargo, hay problemas financieros que limitan el grado de inferencia a grandes regiones, lo que no permite obtener directamente datos desagregados para provincias o distritos, como lo exige la programación y evaluación de los servicios de planificación familiar.

En tanto no se defina un sistema nacional de recopilación continua de información sobre las acciones de planificación familiar, que

considere la cobertura geográfica, la incidencia en la población destinataria, la calidad de los servicios y otras variantes, la única alternativa es realizar estimaciones derivadas de las encuestas especializadas, que proporcionan información confiable aunque limitada a las grandes regiones del país.

El propósito de este estudio es determinar las tasas de prevalencia de uso de anticonceptivos de los departamentos y provincias, así como cuantificar los subgrupos de mujeres consideradas como usuarias o usuarias potenciales de métodos de planificación familiar, datos útiles para la programación y evaluación de los servicios de planificación familiar, que contribuirían también al conocimiento de los cambios ocurridos en los últimos años en el nivel de fecundidad de estos subgrupos de población.

Con este objeto, se aplicará una metodología que combina datos de la ENDES II, de alcance nacional y regional, con datos del Censo Nacional de Población y Vivienda de 1993 correspondientes a departamentos y provincias.

El procedimiento metodológico consta de los siguientes pasos:

- i) Definición de un modelo estadístico que combine dos fuentes de datos (encuesta y censo) para la estimación de tasas en áreas pequeñas;
- ii) Justificación de las variables incluidas, dentro de las limitaciones que conlleva la combinación de dos fuentes de información: ENDES II y el censo de 1993;
- iii) Aplicación de técnicas de análisis multivariado, para determinar la importancia de estas variables con relación a la práctica anticonceptiva, con datos de la ENDES 1991-1992;
- iv) Tabulación de las variables seleccionadas y sus respectivas categorías, a nivel de provincia, a partir del Censo Nacional de Población y Vivienda de 1993, para utilizarlas como ponderadores de la estimación;
- v) Estimación de cobertura de la planificación familiar para las mismas variables y categorías, a partir de datos de la ENDES II correspondientes a cada región;
- vi) Combinación de las dos fuentes de información para hacer estimaciones por departamentos y provincias;
- vii) Cálculo de la precisión de las estimaciones.

I. COMENTARIOS GENERALES

1. Demanda

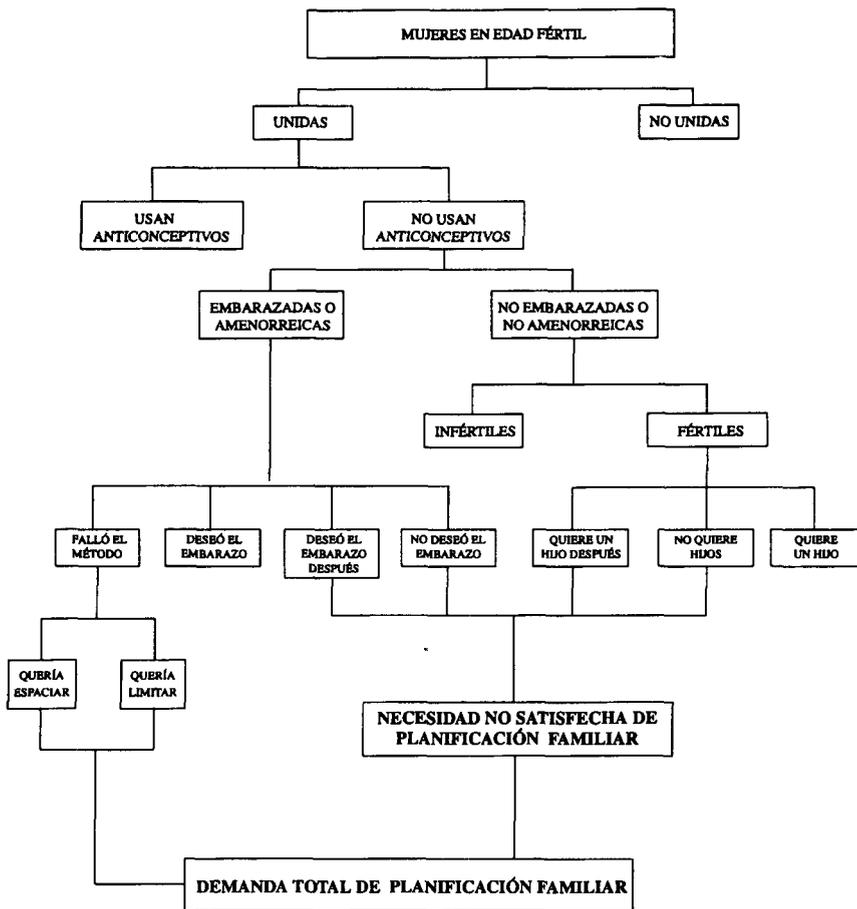
Es evidente que la demanda de planificación familiar está estrechamente vinculada al deseo de las mujeres de tener o no tener más hijos. En el primer caso tendrán que optar por el uso de algún método anticonceptivo. Con la finalidad de estudiar la demanda, en la ENDES II se consideran dos módulos destinados, respectivamente, a determinar las preferencias de la fecundidad y, el conocimiento y uso de anticonceptivos. La combinación de las respuestas recibidas en estos dos módulos permitirá conocer la demanda de planificación familiar y sus componentes.

Para determinar la demanda real de métodos anticonceptivos habría que estudiar a la población total, es decir a todas las mujeres, independientemente de su estado civil. Sin embargo, en el presente estudio se toma en cuenta únicamente a las mujeres casadas o convivientes, por considerarlas más expuestas al riesgo de embarazo. Se excluye a las mujeres que no están en unión conyugal, porque se supone que el riesgo al embarazo es menor; además, hay algunas consideraciones de tipo operativo o práctico que dificultan el preguntar a las mujeres no casadas ni convivientes sobre su actividad sexual o sus intenciones reproductivas.

De acuerdo al planteamiento de Charles Westoff y L.H. Ochoa (1991), las categorías de usuarias de métodos de planificación familiar son: i) mujeres que actualmente están usando anticonceptivos, ii) mujeres con necesidad no satisfecha de uso de anticonceptivos y iii) mujeres embarazadas o amenorreicas, que quedaron embarazadas por falla del método que estaban usando. Para cuantificar estos grupos de mujeres, Westoff propone una metodología, aplicable a las Encuestas de Demografía y Salud (EDS) que se aplica en el presente trabajo (véase la figura 1).

La demanda se clasifica, según las intenciones, en demanda con fines de espaciamiento y con fines de limitación. Esta última corresponde prácticamente al triple de la primera, lo que indica que la mayoría de las mujeres casadas o convivientes no desean tener más hijos (Loza y Vallenás, 1992).

DIAGRAMA 1
DIAGRAMA DE LA DEMANDA DE PLANIFICACIÓN FAMILIAR



2. Uso actual de anticonceptivos

La demanda de planificación familiar proviene de las mujeres o parejas que están usando anticonceptivos, además de un grupo de no usuarias que desean controlar su fecundidad.

Las mujeres que están usando anticonceptivos lo hacen con dos propósitos: posponer el nacimiento del siguiente hijo, o simplemente no tener más hijos, ya sea porque ya alcanzaron el número ideal de hijos o el tamaño de familia que consideran adecuado.

El grupo de mujeres que no están usando anticonceptivos está integrado por mujeres quienes quieren tener un hijo pronto y por otras, no embarazadas ni amenorreicas, que desean postergar el próximo embarazo o no tener más hijos; por último, están las embarazadas o las que recientemente tuvieron un hijo.

3. Mujeres no embarazadas ni amenorreicas

La categoría de mujeres unidas y no usuarias de anticonceptivos se subdivide en dos categorías: embarazadas o amenorreicas con posterioridad a un parto y ni embarazadas ni amenorreicas. Parte de las mujeres no embarazadas ni amenorreicas son fértiles y, por lo tanto están expuestas al riesgo de embarazo. Las que no son fértiles no necesitan protección anticonceptiva, razón suficiente para excluirlas del cálculo de la demanda.

Se define como infértiles a las mujeres no embarazadas que han vivido en unión durante los cinco últimos años y que no han tenido hijos pese a no usar anticonceptivos. Asimismo, se considera en esta categoría a las mujeres no embarazadas que no tuvieron menstruación durante un período de por lo menos 24 semanas (Westoff toma en cuenta un período de 12 semanas) (Westoff y Ochoa, 1991).

Las mujeres fértiles son clasificadas según sus intenciones reproductivas, en las siguientes categorías: i) “espaciadoras”, las que desean postergar el nacimiento de su próximo hijo por lo menos dos años; ii) “limitadoras”, que declaran no desear más hijos, y iii) las que quieren tener un hijo pronto. Este último grupo de mujeres, por razones obvias, es excluido del cálculo de la demanda de métodos anticonceptivos. Se considera que las primeras dos tienen una necesidad insatisfecha de planificación familiar.

4. Mujeres embarazadas o amenorreicas post-parto

Las mujeres embarazadas evidentemente no están expuestas al riesgo de embarazo por lo que cuando se las entrevistó no presentaron una necesidad de planificación familiar. Sin embargo, para incluir a algunas de ellas en el grupo de mujeres con demanda insatisfecha, se supone que tal vez en algunos casos no habrían quedado embarazadas si hubiesen usado anticoncepción, lo que habría evitado embarazos no deseados o no planeados.¹

Las embarazadas que respondieron haber deseado el embarazo en el momento en que ocurrió son excluidas del cálculo de la demanda no satisfecha. A las que querían quedar embarazadas pero esto se produjo antes de lo planeado, se las incluye en el grupo de mujeres con necesidad insatisfecha de espaciamiento; a las que no deseaban quedar embarazadas se las considera en el grupo con necesidad insatisfecha de limitación.

Para clasificar a las mujeres embarazadas o amenorreicas dentro de la demanda no satisfecha se utiliza el concepto de embarazo planeado, a diferencia del grupo de mujeres fértiles, en cuyo caso se consideran las intenciones reproductivas.

5. Falla del método anticonceptivo

Hay una proporción menor de mujeres embarazadas o amenorreicas que manifestaron haber quedado embarazadas debido a una falla del método anticonceptivo que usaban. Estas mujeres se consideran usuarias de métodos anticonceptivos, y pueden clasificarse en “espaciadoras” o “limitadoras”. Este grupo de mujeres no se excluyen en el cálculo de necesidad insatisfecha y pasan a formar parte de la demanda total.

En síntesis, en la medición de la necesidad insatisfecha de planificación familiar se incluye a las mujeres en edad fértil casadas o convivientes, las no embarazadas ni amenorreicas consideradas fértiles, las que no estaban usando un método anticonceptivo y que manifestaron sus deseos de postergar por dos o más años su próximo embarazo, y las que definitivamente no desean tener más hijos. Además, se incluye a las

¹ Para clasificar a las mujeres embarazadas o amenorreicas dentro de la demanda no satisfecha se utiliza el concepto de embarazo planeado, a diferencia del grupo de mujeres fértiles, en cuyo caso se consideran las intenciones reproductivas.

embarazadas o amenorreicas cuyo embarazo no fue planificado en el momento en que ocurrió o que fue un embarazo no deseado.

La demanda total de planificación familiar está conformada por mujeres con necesidad insatisfecha, mujeres que tuvieron un embarazo involuntario por falla del método anticonceptivo y las mujeres en edad fértil unidas que declararon estar usando anticonceptivos.

6. Resultados del estudio

El número de mujeres en edad fértil empadronadas en el censo de 1993 fue de 5 630 378; de este total, 3 250 642 (57.7%) declararon estar casadas o ser convivientes, cifra sobre la cual se estima la demanda total de planificación familiar.

De acuerdo a la ENDES II el 17.1% de la mujeres unidas, casadas o convivientes (556 177) presentan demanda insatisfecha; 1 834 357 (56.4%) mujeres unidas son usuarias de métodos anticonceptivos; 65 622 están expuestas a la falla del método que están utilizando y, por consiguiente, presentarían una demanda de métodos más seguros. La demanda total estaría constituida por 2 456 156 mujeres, el 75.6% del total de casadas o convivientes y el 43.6% de las mujeres en edad fértil.

II. METODOLOGÍA DE ESTIMACIÓN COMBINADA

El presente trabajo tiene por finalidad estimar el número de mujeres unidas que son usuarias de métodos de planificación familiar, por provincias. Con tal objeto se utilizan dos fuentes de información: datos provenientes de la Encuesta Demográfica y de Salud Familiar realizada en 1991-1992 (ENDES II) y del noveno Censo Nacional de Población y Vivienda realizado el 11 de julio de 1993 (censo de 1993).

La utilización de dos fuentes de datos como las mencionadas, debe respaldarse con una metodología que combine la información de ambas. Con tal fin se utilizará el método propuesto por J. García Núñez, Y. Palma C., L. Núñez y J. Jiménez en el trabajo titulado "Estimaciones programáticas para áreas geográficas pequeñas",² que permite correlacionar, mediante una combinación lineal, los datos regionales de

2 José García Núñez (Equipo de Apoyo Técnico del FNUAP); Yolanda Palma y Javier Jiménez, Consejo Nacional de Población, México (CONAPO); Leopoldo Núñez (Pathfinder International).

la ENDES II con los del censo de 1993 desagregados por provincias, para hacer estimaciones provinciales de variables regionales estudiadas en la ENDES II. Los datos regionales, provenientes de la encuesta constituyen la variable dependiente, en tanto que los del censo son las variables independientes.

La ENDES II permite clasificar a las mujeres con relación al uso de anticonceptivos, en las siguientes categorías: usuarias y no usuarias actuales; usuarias esporádicas y mujeres que no han usado nunca métodos de planificación familiar; éstas representan la variable dependiente, con una cobertura que abarca las trece regiones político-administrativas del país. La información proveniente del censo de 1993 proporciona datos sobre frecuencia, por provincias, correspondientes a las variables definidas como independientes o explicativas del uso de métodos de planificación familiar.

Para la selección de las variables independientes se aplica el análisis discriminante a los datos regionales recopilados en la ENDES II. Para ello, se selecciona previamente un conjunto de variables que podrían relacionarse con el uso de métodos de planificación familiar, con la condición fundamental de que estén consideradas tanto en la ENDES II como en el censo de 1993.

Cabe recordar que el análisis discriminante es una técnica del análisis estadístico multivariado, en el que se usan combinaciones lineales de las variables independientes para clasificar los casos individuales en grupos previamente definidos, de manera que muestren homogeneidad dentro del grupo y heterogeneidad entre grupos. La técnica permite, además, identificar las variables que contribuyen en mayor medida a explicar la variable dependiente.

Dado que los datos de ENDES II se agrupan por regiones, el análisis se realiza a ese nivel. Sin embargo, el procedimiento se ilustra con datos nacionales, dejando claramente establecido que las estimaciones finales se realizan tomando como universo cada región.

1. Identificación de las subclases independientes

El primer paso del estudio consiste en identificar “j” subclases en la ENDES II, cada una de las cuales está conformada por grupos homogéneos de mujeres unidas en edad fértil que reúnen las características que definen las variables independientes seleccionadas. Estas variables se escogen en orden de importancia, sobre la base de un

análisis multivariado, cuyo objetivo es explicar las diferencias entre usuarias y no usuarias de anticonceptivos.

A modo de ilustración se describe el ejercicio realizado a nivel nacional, con las siguientes variables independientes seleccionadas mediante el análisis discriminante: servicio de electricidad en la vivienda (QH21A) y nivel de educación (EDUC). En función de estas variables se definen las siguientes subclases de mujeres:

Cuadro 1
SUBCLASES INDEPENDIENTES

Nivel de educación	Electricidad en la vivienda	
	Tiene	No tiene
Sin educación	Subclase 1	Subclase 2
Primaria incompleta	Subclase 3	Subclase 4
Primaria completa
Secundaria
Superior no universitaria
Superior universitaria	Subclase <i>j-1</i>	Subclase <i>j</i>

Fuente: Instituto Nacional de Estadística e Informática (INEI), Encuesta Nacional Demográfica y de Salud Familiar, 1991-1992, Lima, 1992.

2. Estimación muestral de la variable dependiente

Como ya se mencionó, la variable dependiente es el uso de métodos anticonceptivos, que permite clasificar a las mujeres en dos grupos: usuarias y no usuarias. En consecuencia, se deben cuantificar la variable dependiente que se desea estimar a nivel provincial en cada región (*i*) y cada subclase (*j*). Con tal objeto usamos la tasa de prevalencia de uso de métodos anticonceptivos, definida como el cociente entre las usuarias (*x*) y el total de MUEF (*n*). La expresión matemática es:

$$X_{ij} = \frac{x_{ij}}{n_{ij}}$$

En el cuadro 2 se presentan las estimaciones obtenidas de la muestra para las regiones:

Cuadro 2
**CÁLCULO DE LAS ESTIMACIONES DE LA MUESTRA
 PARA REGIONES Y SUBCLASES**

Región	Subclase				
	1	2	...	J = 12	Total
1	$\frac{x_{21}}{n_{21}}$	$\frac{x_{22}}{n_{22}}$...	$\frac{x_{2j}}{n_{2j}}$	$x_{2.} = \frac{\sum x_{2j}}{\sum n_{2j}}$
2	$\frac{x_{31}}{n_{31}}$	$\frac{x_{32}}{n_{32}}$...	$\frac{x_{3j}}{n_{3j}}$	$x_{3.} = \frac{\sum x_{3j}}{\sum n_{3j}}$
3	$\frac{x_{41}}{n_{41}}$	$\frac{x_{42}}{n_{42}}$...	$\frac{x_{4j}}{n_{4j}}$	$x_{4.} = \frac{\sum x_{4j}}{\sum n_{4j}}$
Total	$x_{.1} = \frac{\sum x_{i1}}{\sum n_{i2}}$	$x_{.2} = \frac{\sum x_{i2}}{\sum n_{i2}}$...	$x_{.j} = \frac{\sum x_{ij}}{\sum n_{ij}}$	$x_{..} = \frac{\sum x_{ij}}{\sum n_{ij}}$

Fuente: Instituto Nacional de Estadística e Informática (INEI), Encuesta Nacional Demográfica y de Salud Familiar, 1991-1992, Lima, 1992.

Si el ejercicio se amplía a nivel nacional, las tasas de prevalencia calculadas sobre la base de la encuesta serían las siguientes:

Cuadro 3
TASAS DE PREVALENCIA

Nivel de educación	Electricidad en la vivienda	
	Tiene	No tiene
Sin educación	Tasa de prevalencia 1	Tasa de prevalencia 2
Primaria incompleta	Tasa de prevalencia 3	Tasa de prevalencia 4
Primaria completa
Secundaria
Superior no universitaria
Superior universitaria	Tasa de prevalencia j-1	Tasa de prevalencia j

Fuente: Instituto Nacional de Estadística e Informática (INEI), Encuesta Nacional Demográfica y de Salud Familiar, 1991-1992, Lima, 1992.

3. Ponderación de las subclases

Cada subclase regional calculada sobre la base de la encuesta se pondera por frecuencias relativas correspondientes a cada subclase de las provincias, que a su vez se calculan con datos de las variables independientes provenientes del censo.

Cuadro 4
PONDERACIÓN DE LAS SUBCLASES

Región	Provincia	Subclase				Total
		1	2	...	J	
1	1	W_{11}	W_{12}	...	W_{1j}	$W_{1.}=1$
	2	W_{21}	W_{22}	...	W_{2j}	$W_{2.}=1$
	N	W_{n1}	W_{n2}	...	W_{nj}	$W_{n.}=1$

Fuente: Instituto Nacional de Estadística e Informática (INEI), Encuesta Nacional Demográfica y de Salud Familiar, 1991-1992, Lima, 1992.

La definición de la ponderación “W” para cada provincia “i” y subclase “j” está dada por la siguiente ecuación:

$$W_{ij} = \frac{N_{ij}}{\sum_{j=1} N_{ij}}$$

en la cual se cumple que:

$$\sum_{j=1} W_{ij} = 1$$

Se muestran las ponderaciones a nivel nacional para cada categoría o subclase, que corresponden a la proporción de mujeres unidas en cada categoría, calculada a partir de los datos censales.

Cuadro 5
PONDERADORES

Nivel de educación	Electricidad en la vivienda	
	Tiene	No tiene
Sin educación	Proporción MUEF 1	Proporción MUEF 2
Primaria incompleta	Proporción MUEF 3	Proporción MUEF 4
Primaria completa
Secundaria
Superior no universitaria
Superior universitaria	Proporción MUEF j-1	Proporción MUEF j

Fuente: Instituto Nacional de Estadística e Informática (INEI), Encuesta Nacional Demográfica y de Salud Familiar, 1991-1992, Lima, 1992.

4. Combinación lineal

Una vez calculada la tasa de prevalencia muestral por regiones y las ponderaciones provinciales de cada subclase se procede a una estimación combinada para cada provincia en la que se utiliza la siguiente ecuación:

$$X_i^* = \sum_{j=1} X_j W_j$$

donde X_j es la estimación de cada subclase de la región basada en los datos de la encuesta y W_j es la ponderación de cada subclase y provincia basada en los datos censales.

La tasa estimada de prevalencia nacional sería la siguiente:

$$X^* = X_1 * W_1 + X_2 * W_2 + \dots + X_j * W_j$$

5. Estimación del número de usuarias de anticonceptivos

Una vez estimada la tasa de prevalencia provincial (paso anterior), se multiplica por el número de mujeres unidas de cada provincia, dato proporcionado por el censo. Esto permite calcular el total de mujeres usuarias de métodos de planificación familiar en cada provincia.

$$M_i = N_i * X_i^*$$

La suma de las usuarias de cada provincia da el total departamental respectivo que dividido por MUEF, da la estimación de

la tasa de prevalencia departamental. El mismo procedimiento se emplea para calcular la prevalencia regional estimada.

III. LAS VARIABLES EN ESTUDIO

A partir de la década de 1970, se han realizado en Perú varias encuestas nacionales con el objeto de recopilar los datos necesarios para conocer el comportamiento de las mujeres con respecto a su fecundidad y las prácticas anticonceptivas. Las más conocidas son la encuesta de fecundidad urbana y rural de 1969-1970, la encuesta nacional de fecundidad del Perú de 1977-1978, encuesta nacional de prevalencia y uso de anticonceptivos de 1981, y la encuesta demográfica y de salud familiar de 1986 y 1991-1992 (ENDES I y ENDES II). También se han realizado actividades destinadas a estudiar algunos de los factores o variables intermedias relacionadas con las decisiones sobre la fecundidad y su regulación sobre la base de los resultados de estas encuestas (Mostajo, 1981; Ortiz y Alcántara, 1988). Sin embargo, muy poco se sabe sobre los cambios de la fecundidad y su relación con las prácticas anticonceptivas en grupos de mujeres de áreas geográficas pequeñas, provincias o distritos, debido indudablemente al limitado alcance de los datos provenientes de las encuestas que, en el mejor de los casos, presentan cifras correspondientes a grandes regiones (ENDES II).

El propósito de este trabajo es contribuir al conocimiento de las prácticas anticonceptivas en las provincias, mediante estimaciones de tasas de prevalencia de uso de anticonceptivos a esos niveles, a partir de los datos regionales que proporciona la ENDES II.

José García Núñez y otros sostienen que para comprender el proceso que conduce a la decisión de usar un método anticonceptivo, hay que considerar tres elementos básicos que parecen seguir una secuencia: en primer lugar, debe existir una percepción de la posibilidad de limitar el tamaño de la familia; en segundo término, deben haber elementos que motiven a una pareja, ya sea para tener menos hijos o para espaciar sus nacimientos; por último, la decisión de adoptar un método anticonceptivo debe ser el resultado de la evaluación positiva de por lo menos un método y de una fuente, en el caso de los métodos no tradicionales (García Núñez y otros 1994).

Dada la limitación fundamental del método de estimación, consistente en utilizar variables explicativas que se hayan incluido tanto en la ENDES II como en el censo de 1993, se consideran únicamente como variables independientes las referidas a algunos aspectos

socioeconómicos y demográficos de la mujer, que definirían el contexto de la decisión de recurrir a la anticoncepción; se deja de lado otras variables intermedias de gran importancia, que inciden sobre todo en la motivación para regular la fecundidad, como: el tamaño ideal de la familia o fecundidad ideal, el deseo de espaciar o limitar el número de hijos y el conocimiento de por lo menos un método y una fuente, especialmente en el caso de los métodos no tradicionales, que estarían estrechamente vinculados a la decisión de adoptar una práctica anticonceptiva. Dado que las estimaciones se realizan para las trece regiones geográficas que conforman el país, se estaría incorporando implícitamente la dimensión cultural, que explicaría algunas diferencias regionales en la selección del tipo de variable independiente.

1. Variables de la encuesta y el censo

Para facilitar la selección de las variables incluidas en el estudio, se presenta la relación entre las estudiadas en la encuesta y en el censo.

Características de las mujeres

VARIABLE	ENDES II	CENSO 1993
- Elegibilidad	sí	no
- Relación con el jefe del hogar	sí	sí
- Residencia habitual en la vivienda	sí	no
- Sexo	sí	sí
- Edad	sí	sí
- Lugar de nacimiento	no	sí
- Lugar de residencia habitual	sí	sí
- Supervivencia de la madre	sí	sí
- Idioma o dialecto materno	sí	sí
- Alfabetismo	sí	sí
- Nivel de instrucción	sí	sí
- Asistencia escolar	sí	sí
- Profesión u oficio	sí	sí
- Actividad económica	sí	sí
- Categoría de ocupación	no	sí
- Estado civil	sí	sí
- Religión	sí	sí
- Hijos nacidos vivos y sobrevivientes	sí	sí
- Fecha de nacimiento del último hijo	sí	sí
- Supervivencia del último hijo	sí	sí
- Supervivencia de hermanas	sí	no
- Preguntas sobre embarazo, parto y menstruación	sí	no
- Conocimiento y uso de anticonceptivos	sí	no
- Preferencias de fecundidad	sí	no
- Acceso al Seguro Social	sí	no

Características y servicios de la vivienda

VARIABLE	ENDES II	CENSO 1993
- Tipo de vivienda	no	sí
- Condición de ocupación de la vivienda	no	sí
- Material predominante en las paredes	no	sí
- Material predominante en el techo	no	sí
- Material predominante en el piso	no	sí
- Abastecimiento de agua	sí	sí
- Tipo de servicio higiénico	sí	sí
- Alumbrado eléctrico	sí	sí
- Número de habitaciones	no	sí
- Equipamiento del hogar	sí	sí
- Tenencia de medios de comunicación	sí	sí

2. Selección de variables y definición de categorías

a) Estado civil de la mujer

En los estudios de la fecundidad, una variable que se considera determinante es la nupcialidad, puesto que ésta muestra la proporción de mujeres en edad reproductiva que tienen relaciones sexuales estables. Esto las sitúa como mujeres más expuestas que otras al riesgo de embarazo, razón que explica que en el grupo de mujeres casadas o convivientes haya un mayor número de nacimientos.

Es evidente que la exposición al riesgo de embarazo, definida por el estado civil de la mujer, es la principal condicionante para el uso de métodos anticonceptivos. Los resultados de la ENDES II (véase el cuadro 6) muestran que la gran mayoría de las mujeres en edad fértil (cerca de las dos terceras partes) no usa métodos anticonceptivos, proporción que aumenta en el caso de las mujeres solteras, puesto que un 95% de ellas no usa métodos anticonceptivos. Sin embargo, en el grupo de mujeres casadas y convivientes, las proporciones de no usuarias actuales de métodos anticonceptivos están por debajo de la mitad. Al analizar las proporciones de mujeres que alguna vez fueron usuarias de métodos anticonceptivos y las que nunca lo fueron, el sentido de los diferenciales es el mismo, aunque las diferencias se agudizan más.

El estado civil de las mujeres se ha estudiado tanto en la encuesta como en el censo, sin embargo, éste no se considera como variable independiente debido a que está estrechamente ligado a la práctica anticonceptiva, pues antes que explicar el uso de anticonceptivos, lo

Cuadro 6
**CONDICIÓN DE USO DE MÉTODOS ANTICONCEPTIVOS DE LAS MUJERES
 EN EDAD FÉRTIL, SEGÚN ESTADO CIVIL, 1992**

Estado civil	Condición de uso de anticonceptivos (porcentajes)			
	Uso actual	No uso actual	Usó alguna vez	Nunca usó
Total	35.63	64.32	54.57	45.41
Soltera	5.08	94.92	11.04	88.94
Unida	58.97	41.03	82.76	17.23
Otro	18.73	81.18	70.57	29.43

Fuente: Instituto Nacional de Estadística e Informática (INEI), Encuesta Nacional Demográfica y de Salud Familiar, 1991-1992, Lima, 1992.

estaría condicionando. Por esta razón, el análisis se restringe al grupo de mujeres en edad fértil casadas o convivientes (unidas maritalmente) al momento de la entrevista.

b) Nivel de educación de la mujer

Durante los últimos años ha habido una importante disminución de las tasas de fecundidad en Perú. Muchas investigaciones confirman que esta disminución sería producto de la creciente urbanización de la sociedad y, principalmente, del aumento de los niveles de educación, en desmedro de la incidencia que podría tener el aborto. El aumento de la participación de la mujer en el campo profesional y laboral la ha obligado a dedicar más tiempo a estas actividades, lo que no sólo significa una reorientación de su actitud frente a los distintos roles que le toca desempeñar en su vida, sino también un nuevo modo de percibir y conceptualizar la maternidad que adquiere diferentes matices según el nivel educacional alcanzado.

El nivel de educación es una variable que en alguna medida define una posición social de la mujer que le permite visualizar motivaciones diferenciales respecto a la fecundidad y su regulación. Esto se debe, entre otras razones, a que la educación no sólo cambia el planteamiento de la mujer con respecto a los hijos, sino que también posibilita el acceso a la información acerca de métodos de planificación familiar y la capacidad para usarlos. En las mujeres de mayor nivel de educación,

la valoración de los hijos y el número de éstos están determinados por la consecución de metas trazadas o bien como una manera de realización personal, e indudablemente tiene alta relación con la preferencia sobre el tamaño de la familia; en cambio, en los grupos de mujeres con menor nivel de educación, sobre todo las que viven en áreas rurales, los hijos se valoran en términos de la ayuda, el apoyo económico y la compañía que puedan proporcionar. En algunos casos incluso se piensa en los hijos como un amparo para la vejez. Estos factores permiten asociar altos niveles de fecundidad y bajo uso de anticonceptivos.

Los resultados de las ENDES y otras encuestas y censos permiten señalar que el nivel de educación es una variable que está íntimamente ligada con las tasas de fecundidad. Según la ENDES II, la tasa global de fecundidad (TGF) en las mujeres sin educación es de 7.4 hijos en promedio, mientras que las que han cursado hasta el nivel superior de educación tienen 2.7, es decir, casi 5 hijos de diferencia, y las mujeres que declararon uno o varios años de primaria aprobados tenían una TGF de 6.1, cifra bastante cercana a la de las que se declaran sin educación. El promedio baja drásticamente a 3.8 hijos en las mujeres que tienen uno o varios años de secundaria. Este cambio brusco resalta la necesidad de definir un grupo intermedio entre las mujeres de nivel primario de educación, a fin de graficar mejor los diferenciales según esta variable.

Similar apreciación se puede realizar en cuanto al uso de métodos anticonceptivos. En efecto, Nelly Mostajo (1981) concluye que “la educación de la mujer es la variable más importante en la explicación del uso actual de métodos anticonceptivos, marcando el nivel de primaria completa en el área rural y secundaria o más en el área urbana como los puntos de corte donde se observa un cambio significativo de usuarias”. El uso de métodos anticonceptivos asociado al nivel de educación de la mujer es un indicador importante con respecto al modo en que las parejas tratan el tema de la limitación de la fecundidad, según la ENDES II, el mayor diferencial en el uso actual de anticonceptivos se observa entre las mujeres sin educación y las que cursaron uno o varios años de primaria, con 35.1% y 52.1% de prevalencia en el uso de éstos, respectivamente (Loza y Vallenas, 1992). Por otro lado, la prevalencia más alta se da entre las mujeres con algún grado de educación superior universitaria (75.2%). También se observa que más del 90% de las mujeres en edad fértil que finalizaron

Cuadro 7
**CONDICIÓN DE USO DE MÉTODOS ANTICONCEPTIVOS
 DE LAS MUJERES UNIDAS EN EDAD FÉRTIL,
 SEGÚN ESCOLARIDAD, 1991-1992**

Escolaridad	Uso actual (%)	Usó alguna vez (%)
Sin educación	35.12	56.40
Primaria incompleta	46.68	72.67
Primaria completa	57.18	81.41
Secundaria	65.63	90.35
Superior no universitaria	70.85	93.22
Superior universitaria	75.15	94.93
Total	58.98	82.77

Fuente: Instituto Nacional de Estadística e Informática (INEI), Encuesta Nacional Demográfica y de Salud Familiar, 1991-1992, Lima, 1992.

la secundaria o tienen nivel superior recurrieron, en algún momento de su vida, al uso de algún método anticonceptivo, en tanto que las mujeres sin educación están cercanas al 50% de prevalencia en el uso de métodos anticonceptivos. La proporción de mujeres “nunca usuarias” es más contundente aún, puesto que poco menos de la mitad de las mujeres sin educación pertenecen a este grupo, mientras que en las que tienen nivel superior de educación la proporción sólo alcanza al 5%.

Estas cifras comprueban la afirmación que señala que a mayor tiempo de permanencia dentro del sistema escolar o mayor cantidad de estudios destinados a incorporarse al sistema laboral, la preocupación por controlar la fecundidad también es mayor.

Para fines del presente estudio, los niveles de educación se articularon tomando en cuenta los resultados mencionados. Inicialmente, se distinguieron cuatro grupos: i) sin educación, ii) primaria, iii) secundaria y, iv) superior. Sin embargo, fue necesario separar el nivel de educación primaria en dos subgrupos: i) incompleta (cuando se declara haber cursado de 1 a 4 años de estudio) y ii) completa (5 ó 6 años de estudio); y el nivel de educación superior en: i) no universitaria (algún año de educación superior no universitaria) y, ii) superior universitaria.

En consecuencia se proponen las seis categorías siguientes:

- i) Sin educación: abarca a las mujeres entrevistadas que declararon no haber cursado o aprobado ningún año de estudios en el sistema educativo convencional, o las que declararon haber cursado el nivel primario y no aprobaron ningún año (o grado). Se agrega, además, la categoría de “no especificado”, cuyo número de casos es pequeño.
- ii) Primaria incompleta: incluye a las mujeres que declararon haber aprobado de 1 a 4 años (o grados) del nivel primario.
- iii) Primaria completa: incluye a las mujeres que aprobaron 5 ó 6 años (o grados) del nivel primario. Se incluye, además, a las que declararon poseer nivel secundario y no aprobaron ningún año (o grado).
- iv) Secundaria: incluye a las mujeres que aprobaron algún año (o grado) del nivel secundario y a las que declararon poseer nivel superior y no aprobaron ningún año.
- v) Superior no universitaria: considera a las mujeres que declararon haber aprobado algún año de educación superior no universitaria.
- vi) Superior universitaria: considera a las mujeres que declararon haber aprobado algún año de educación superior universitaria.

c) Área de residencia

El área de residencia de una persona define un contexto económico cultural específico que permite diferenciar tanto las actitudes frente a la fecundidad como la práctica anticonceptiva. En Perú existe una marcada desigualdad en la cobertura de los servicios de salud en el territorio. Esta situación significa que el acceso a los servicios de planificación familiar -los cuales están integrados a los servicios formales de salud- sea bastante distinto en las localidades urbanas y rurales. Por otro lado, al margen del nivel de educación y de la disponibilidad de servicios de salud, la organización de los hogares en cada área de residencia define la forma en que se valora a los hijos. Éste es el caso de las áreas rurales, donde a menudo se espera de los hijos un apoyo en el trabajo agrícola y doméstico, lo cual influye de manera decisiva en la regulación de la fecundidad. A esto se debe agregar que los costos asociados a la práctica anticonceptiva (métodos no tradicionales) generalmente son más altos en las áreas que están alejadas de los centros urbanos.

Dos tercios de las mujeres residentes en áreas urbanas usaban anticonceptivos al realizarse la ENDES II. Por otra parte, también dentro del área urbana, el 90% de las mujeres alguna vez usó un método anticonceptivo, a diferencia del 65% de mujeres residentes en zonas rurales.

Es importante señalar que la discriminación entre área urbana y rural oculta diferencias bastante marcadas dentro de cada una de ellas con relación al tamaño de las localidades. La definición de área rural que se utilizó para la ENDES II se refiere a localidades con menos de 2 000 habitantes, las localidades con población superior engloban a las áreas urbanas. Está claro que se encontrarían diferencias importantes si se estudiaran, por ejemplo áreas marginales o localidades de 2 000 a 20 000 y de 20 000 a 100 000, y así sucesivamente, y diferencias aún mayores si se establecieran distinciones entre zonas costeras, selváticas y serranas.

Sin embargo, el manejo de los datos censales se verá dificultado por estos criterios cuantitativos de área de residencia, puesto que en el censo se emplean definiciones de tipo operativo. En todo caso, se asume el riesgo de tomar grupos poblacionales no estrictamente comparables aunque se supone que estos poseen características comunes.

Cuadro 8
CONDICIÓN DE USO DE MÉTODOS ANTICONCEPTIVOS DE
LAS MUJERES EN EDAD FÉRTIL UNIDAS, SEGÚN ÁREA
DE RESIDENCIA, 1991-1992

Área de residencia	Condición de uso de anticonceptivos (%)	
	Uso actual	Usó alguna vez
Total	58.98	82.77
Urbano	66.15	89.65
Rural	41.12	65.63

Fuente: Instituto Nacional de Estadística e Informática (INEI), Encuesta Nacional Demográfica y de Salud Familiar, 1991-1992, Lima, 1992.

d) Características de la vivienda

Ante la ausencia de información sobre el nivel de ingreso de los hogares, las características de la vivienda permiten definir un contexto

socioeconómico para realizar estudios diferenciales, debido a que el tipo y las condiciones de la vivienda son el resultado de la capacidad económica de los hogares y de la disponibilidad de servicios comunitarios, factores que en gran medida definen su forma de vida. Recientes estudios sobre las carencias de las viviendas permiten deducir un estrecho vínculo entre éstas y los grados de pobreza de la población, por lo que algunas características de la vivienda están íntimamente relacionadas con el comportamiento reproductivo en particular y, por ende, con el uso de anticonceptivos.

José García y otros (1994) refiriéndose a las características de la vivienda, sostienen que “estas variables no ejercen necesariamente su influencia a través de las dimensiones que antes se han establecido: la percepción de la posibilidad de limitar el tamaño de la familia, la motivación de tener menos hijos o los costos de regular la fecundidad, sino más bien porque están altamente asociadas con muchas de las variables independientes socioeconómicas que se han considerado, por ejemplo, la educación, el ingreso y el tamaño de la localidad de residencia”. Debido a la disponibilidad de estas variables a nivel de provincia en el censo, es de suma utilidad tomarlas en cuenta si se considera que el método para realizar estimaciones en áreas pequeñas debe utilizar el menor número posible de variables.

De las características escogidas para determinar las condiciones de vivienda, se puede observar que todas aquellas que tienen relación con un mejor servicio o que pueden significar un mejor nivel de vida (electricidad, red pública como fuente de agua, material de piso de un material que no sea tierra, red pública como desagüe) no sólo alcanzan porcentajes más altos en el uso actual de anticonceptivos, sino también dentro del grupo de mujeres alguna vez usuarias de métodos anticonceptivos (véase el cuadro 8). Esta constatación empírica indica una estrecha relación entre las buenas condiciones de la vivienda y una mejor comprensión de las necesidades de limitar la fecundidad a través del uso de anticonceptivos, relación que estaría funcionando a través de otras variables intermedias, como se señaló anteriormente.

Por otro lado, la disponibilidad de electricidad permite el acceso a los medios de comunicación, factor que podría tener una fuerte incidencia en el conocimiento sobre la regulación de la fecundidad, en cuanto a sus costos y mejor utilización. Sobre todo tomando en cuenta el hecho de que en los últimos años, gracias a los programas de política de población fomentados por el gobierno, se implementaron campañas

Cuadro 9

CONDICIÓN DE USO DE MÉTODOS ANTICONCEPTIVOS DE LAS MUJERES EN EDAD FÉRTIL UNIDAS, SEGÚN CARACTERÍSTICAS DE LA VIVIENDA

Características de la vivienda	Uso de anticonceptivos (porcentajes)	
	Uso actual	Usó en el pasado
Electricidad		
Tiene	66.53	89.79
No tiene	41.51	66.54
Agua en la vivienda		
Red pública	65.11	88.86
Otro	44.58	68.48
Desagüe		
Red pública	70.09	92.80
Otro	48.34	73.17
Material de pisos		
Tierra	45.34	70.49
Otros	67.08	90.06

Fuente: Instituto Nacional de Estadística e Informática (INEI), Encuesta Nacional Demográfica y de Salud Familiar, 1991-1992, Lima, 1992.

de información sobre planificación familiar. Otras características, como por ejemplo el acceso a los servicios de agua y desagüe conectados a red pública, y el tipo de material predominante en la construcción de la vivienda, están más ligadas a la salud de la familia y son, en alguna medida, indicadores del nivel de vida de la población, sobre todo en los conglomerados urbanos.

La condición de la vivienda se determina sobre la base de cuatro criterios: i) disponibilidad de servicios de agua potable, ii) disponibilidad de servicio de desagüe, iii) disponibilidad de energía eléctrica y iv) material predominante en los pisos. Las categorías resultantes son las siguientes:

- Agua conectada a red pública: sí, dentro de la vivienda; sí, fuera de la vivienda; no dispone de agua conectada a red pública.
- Desagüe: sí, está conectado a red pública; sí, está conectado a fosa séptica; no tiene desagüe.
- Energía eléctrica (disponible, no disponible).
- Material de los pisos (disponible, no disponible).

e) Edad

Las estructuras por edad y sexo son de importancia fundamental para el análisis de los datos provenientes de encuestas y censos, no solamente en los estudios propios de dichas estructuras, sino también por las diferencias que revelan cuando se las relaciona con otros aspectos de la población, como el estado civil, las características educacionales y socioeconómicas.

En las estadísticas censales o de otro tipo, se puede constatar que el principal efecto de la ampliación de la cobertura de los servicios educacionales en los últimos años es una mayor escolaridad de las mujeres más jóvenes, lo cual estaría generando, entre otras cosas, actitudes “más modernas” en relación con la sexualidad, los roles dentro de la pareja y la sociedad, las expectativas de vida, y otros aspectos que influyen directamente en su comportamiento reproductivo.

Como ya se señaló, el uso de anticonceptivos está estrechamente relacionado con el grado de riesgo de embarazo de una mujer, el cual está definido en gran medida por la edad, factor que tiene gran influencia sobre muchas de las variables intermedias, entre otras el inicio de las uniones maritales, la frecuencia de relaciones sexuales, la esterilidad, la menopausia y el estado civil.

En el cuadro 10 se observa que la edad es una variable asociada al uso de métodos anticonceptivos. Sin embargo, la asociación entre estas variables no es lineal, puesto que el incremento de la edad va acompañado al aumento de la prevalencia del uso de métodos anticonceptivos solamente hasta el rango que va de 35 a 39 años de edad, a partir del cual desciende hasta valores cercanos a los del primer grupo de edades.

Los niveles (tasas) de prevalencia más bajos se presentan en los extremos de las edades de las mujeres en edad fértil, 29% para el quinquenio más joven y 42% para las más adultas. En el primer caso esto se debe seguramente al hecho de no estar posponiendo el nacimiento del primer hijo y en el segundo al incremento de las tasas de esterilidad o bien la disminución del riesgo reproductivo producto de la menor frecuencia coital. Por otro lado, las tasas más altas de uso de anticonceptivos, de 60% a 70%, se presentan en los cuatro grupos que van de los 25 a los 44 años. Dentro del primer grupo etario, es importante destacar la gran cantidad de mujeres que alguna vez han sido usuarias de métodos anticonceptivos. Esto significa que más de la mitad de las mujeres jóvenes de entre 15 y 19 años ya han experimentado con el uso de anticonceptivos, lo que constituye un

Cuadro 10

**USO DE MÉTODOS ANTICONCEPTIVOS DE LAS MUJERES EN EDAD
FÉRTIL UNIDAS SEGÚN GRUPOS QUINQUENALES DE EDAD, 1991-1992**

Edad en grupos quinquenales	Uso de anticonceptivos (porcentajes)	
	Uso actual	Usó en el pasado
15 a 19 años	29.07	54.65
20 a 24 años	49.11	76.56
25 a 29 años	59.50	84.53
30 a 34 años	67.35	88.84
35 a 39 años	69.91	87.72
40 a 44 años	63.77	83.76
45 a 49 años	42.66	77.60

Fuente: Instituto Nacional de Estadística e Informática (INEI), Encuesta Nacional Demográfica y de Salud Familiar, 1991-1992, Lima, 1992.

aumento importante en relación a los datos de la ENDES I (1986), en que el porcentaje era sólo de 41%.

Las categorías de edades utilizadas en la encuesta son los grupos quinquenales a partir de los 15 años.

f) Número de hijos

El resultado de las distintas opciones que ha tomado la mujer a lo largo de su vida reproductiva se refleja en el número de hijos nacidos vivos. Estas opciones están, sin duda, estrechamente vinculadas a la práctica anticonceptiva. Según la ENAF (1976), la proporción de mujeres unidas sin hijos que usaban algún método anticonceptivo fue sólo de 6.7%, porcentaje que se eleva a 16.5% en 1986 (ENDES I) y a 23.4% en 1991-1992 (ENDES II). Estas cifras muestran que en Perú las parejas están posponiendo cada vez más el nacimiento del primer hijo.³ Luego, la proporción de usuarias de métodos anticonceptivos aumenta rápidamente en función del número de hijos tenidos, debido a la gran injerencia de las "limitadoras" (o mujeres que desean controlar la fecundidad) dentro de este grupo. Esto significa que hay un considerable incremento en la motivación por reducir el tamaño de la familia y, además, muestra la evidente relación que hay entre esta motivación y la práctica anticonceptiva.

³ Según los resultados de la ENDES II, solamente el 1.3% de las mujeres unidas sin hijos son limitadoras, proporción que aumenta rápidamente a 16.6% y a 50.5% para las que tienen 1 y 2 hijos respectivamente (Loza y Vallenar, 1992).

Aunque la edad y el número de hijos están estrechamente relacionados, esta última variable parece explicar mejor el uso de anticonceptivos en los primeros quinquenios. Sólo la cuarta parte de las mujeres que no tenían hijos controlaban su fecundidad, en tanto las que ya tenían un hijo llegan a casi el 60%. En las distintas categorías de edad no se ven cambios tan marcados. Por otra parte, la limitación de nacimientos luego del segundo hijo pareciera estar más asociada con el incremento de la edad.

Cuadro 11
**USO DE MÉTODOS ANTICONCEPTIVOS DE LAS MUJERES
 EN EDAD FÉRTIL UNIDAS, SEGÚN NÚMERO
 DE HIJOS TENIDOS, 1991-1992**

Número de hijos tenidos	Condición de uso de anticonceptivos (%)	
	Uso actual	Usó en el pasado
0	23.43	54.04
1	59.75	81.96
2	66.71	88.54
3	65.13	88.64
4	64.49	89.16
5 y más	52.58	77.32
Total	58.98	82.77

Fuente: Instituto Nacional de Estadística e Informática (INEI), Encuesta Nacional Demográfica y de Salud Familiar, 1991-1992, Lima, 1992.

IV. ANÁLISIS MULTIVARIADO

Como ya se ha dicho, el principal objetivo del presente estudio es estimar indicadores de demanda de planificación familiar para “áreas pequeñas” (provincias) mediante la combinación de dos fuentes de información: ENDES II y el censo de población de 1993. Con los datos de la ENDES es posible identificar las variables que determinan el uso de anticonceptivos, tanto a nivel nacional como regional y en los del censo se encuentran las ponderaciones para los departamentos y provincias.

La metodología exige una clasificación de la variable dependiente (uso de métodos anticonceptivos), a partir de un grupo de variables independientes identificadas en el marco conceptual. Para ello, y con el propósito de lograr la mejor clasificación, se probaron varios métodos de análisis estadístico. Por una parte, se utilizará el análisis

discriminante, a través del cual se construye una función, que consiste específicamente en una combinación lineal de las variables independientes, lo que permite predecir la pertenencia a cada uno de los grupos a través del valor de la función generada, que es la función discriminante. Además, se selecciona, a modo de ensayo, la regresión logística, técnica que también genera una ecuación, pero con una aproximación al problema un tanto diferente a la del análisis discriminante, ya que se utiliza un enfoque probabilístico, en el que se estima directamente la probabilidad de que una mujer de determinadas características haga uso de anticonceptivos.

De acuerdo a la experiencia en este tipo de estudios (J. García y otros, 1994) y en base a resultados previos con análisis discriminante, se seleccionó a los siguientes grupos de mujeres como objeto de estudio:

Grupo 1) “Alguna vez usuarias”, “nunca usuarias” de métodos anticonceptivos.

Grupo 2) “Usuarias actuales”, “no usuarias actuales” de métodos anticonceptivos.

Grupo 3) “Usuarias actuales”, “nunca usuarias” de métodos anticonceptivos.

Los porcentajes de casos correctamente clasificados para los grupos en estudio son los siguientes: grupo 1, 71.54%; grupo 2, 63.94% y grupo 3, 73.64%.

No se tomó en consideración el segundo grupo, dado su bajo porcentaje de clasificación correcta. Una vez determinados los subgrupos de mujeres, se profundizó el análisis de la forma en que cada una de las variables seleccionadas influía en la determinación del modelo.

1. Análisis discriminante

a) Nivel nacional

Un indicador elemental que permite una primera aproximación sobre la capacidad de una variable independiente para diferenciar los grupos es la media de la variable de cada uno de los subgrupos estudiados. Mientras mayor es la diferencia de la media en los dos grupos, ésta distingue en mayor medida a las subpoblaciones. Si a esta medida de tendencia central agregamos otra, de dispersión (desviación estándar), tendremos una idea un poco más precisa de la capacidad de la variable para discriminar, puesto que uno de los objetivos del análisis de discriminantes es minimizar la variancia al interior de cada uno de los grupos y maximizar la variancia entre los grupos. A continuación se

presentan las medias y variancias para la clasificación de las mujeres “alguna vez usuarias” y “nunca usuarias”, y para la de “usuarias actuales” y “nunca usuarias” de métodos anticonceptivos, correspondientes a las variables consideradas en el estudio.

Cuadro 12
ALGUNA VEZ USUARIAS Y NUNCA USUARIAS MEDIAS Y DESVIACIONES ESTÁNDAR DE LAS VARIABLES DISCRIMINANTES, POR GRUPO, 1991-1992

Variable	Alguna vez usuarias		Nunca usuarias		Total	
	Media	Desviación	Media	Desviación	Media	Desviación
Agua	0.71	0.45	0.43	0.49	0.67	0.47
Electricidad	0.76	0.43	0.44	0.50	0.71	0.46
Piso	0.33	0.47	0.60	0.49	0.37	0.48
Desagüe	0.50	0.50	0.20	0.40	0.45	0.50
Área de residencia	1.23	0.42	1.54	0.50	1.28	0.45
Número de hijos	3.08	1.52	2.69	1.84	3.02	1.58
PEA	3.87	2.51	4.07	2.46	3.90	2.50
Edad	32.66	7.91	29.84	9.51	32.20	8.26
Educación	2.68	1.20	1.94	1.10	2.56	1.18

Fuente: Instituto Nacional de Estadística e Informática (INEI), Encuesta Nacional Demográfica y de Salud Familiar, 1991-1992, Lima, 1992.

Cuadro 13
USUARIAS ACTUALES Y NUNCA USUARIAS MEDIAS Y DESVIACIONES ESTÁNDAR DE LAS VARIABLES DISCRIMINANTES, POR GRUPO, 1991-1992

Variable	"Usuarias actual"		Total	
	Media	Desviación	Media	Desviación
Agua	0.74	0.44	0.67	0.47
Electricidad	0.78	0.41	0.71	0.45
Piso	0.30	0.46	0.37	0.48
Desagüe	0.53	0.50	0.46	0.50
Área de residencia	1.21	0.41	1.29	0.45
Número de hijos	3.10	1.47	3.0	1.57
PEA	3.79	2.52	3.8	3.00
Edad	32.94	7.47	32.3	8.05
Educación	2.76	1.20	2.6	1.23

Fuente: Instituto Nacional de Estadística e Informática (INEI), Encuesta Nacional Demográfica y de Salud Familiar, 1991-1992, Lima, 1992.

En estos dos cuadros se observa que las variables que muestra menores diferencias en sus medias son la población económicamente activa (PEA) y la edad. Además, las mayores diferencias se observan entre los grupos de “usuarias actuales” y “nunca usuarias”, en los que las mayores disparidades en las medias se dan en las variables relacionadas con la vivienda, la educación y el número de hijos. Al examinar las desviaciones estándar de las variables, no se encontraron variaciones importantes entre los grupos de “usuarias actuales” y “alguna vez usuarias”, aunque sí existen marcadas diferencias en las variables relativas a los servicios de las viviendas.

Un segundo elemento que contribuyó de manera importante a determinar si una variable debía o no incluirse en el modelo discriminante fue la alta correlación existente entre algunas variables, lo que finalmente condujo a la siguiente selección de variables a nivel nacional sobre uso actual:

a) Nivel nacional

Variable	Nombre	Wilks Lambda
1. QH21A	Hogar con electricidad	0.95662
2. EDUC	Nivel de educación de la mujer	0.94239
3. Q208	Total de hijos	0.92699
4. QWURBRUR	Área de residencia (urbana o rural)	0.92216
5. QH16	Servicio de agua en la vivienda	0.92153
6. AGE	Edad	0.92088

b) Nivel regional

Un análisis similar al realizado a nivel nacional nos lleva a concluir que la importancia de las variables independientes varía según la región con la que se trabaje, lo que nos permite la siguiente selección regional sobre uso actual:

Variable	Nombre	Wilks Lambda
Región Loreto		
1. QH16	Servicio de agua en la vivienda	0.86025
2. Q208	Total de hijos	0.82537
3. EDUC	Nivel de educación de la mujer	0.77053
4. QH21A	Electricidad en el hogar	0.75872

Región A.A. Cáceres

1.	QH21A	Electricidad en el hogar	0.96904
2.	EDUC	Nivel de educación de la mujer	0.95945
3.	Q208	Total de hijos	0.94810
4.	QWURBRUR	Área de residencia (urbana o rural)	0.94233

Región Arequipa

1.	QH16	Servicio de agua en la vivienda	0.96379
2.	EDUC	Nivel de educación de la mujer	0.94817
3.	Q208	Total de hijos	0.93917

Región Chavin

1.	QH21A	Electricidad en el hogar	0.95255
2.	QH16	Total de hijos	0.93627
3.	EDUC	Nivel de educación de la mujer	0.91487
4.	QWURBRUR	Área de residencia (urbana o rural)	0.92465

Región Grau

1.	EDUC	Nivel de educación de la mujer	0.96735
2.	Q208	Total de hijos	0.93999
3.	QWURBRUR	Área de residencia (urbana o rural)	0.92540
4.	QH16	Servicio de agua en la vivienda	0.92351

Región Inka

1.	QH21A	Electricidad en el hogar	0.92674
2.	EDUC	Nivel de educación de la mujer	0.91507
3.	Q208	Total de hijos	0.89365
4.	QWURBRUR	Área de residencia (urbana o rural)	0.89152

Región Mariátegui

1.	QH21A	Electricidad en el hogar	0.93571
2.	QWURBRUR	Área de residencia (urbana o rural)	0.91717
3.	Q208	Total de hijos	0.90894
4.	QH16	Servicio de agua en la vivienda	0.90631

Región Los Libertadores

1.	EDUC	Nivel de educación de la mujer	0.91392
2.	Q208	Total de hijos	0.89537
3.	QH21A	Electricidad en el hogar	0.88056
4.	QH16	Servicio de agua en la vivienda	0.87391

Región Nororiental del Marañón

1.	QH21A	Electricidad en el hogar	0.97995
2.	Q208	Total de hijos	0.96501
3.	EDUC	Nivel de educación de la mujer	0.94860

Región La Libertad

1.	QWURBRUR	Área de residencia (urbana o rural)	0.96426
2.	Q208	Total de hijos	0.94500
3.	EDUC	Nivel de educación de la mujer	0.91452

Región San Martín

1.	Q208	Total de hijos	0.97210
2.	EDUC	Nivel de educación de la mujer	0.93881
3.	QWURBRUR	Área de residencia (urbana o rural)	0.92110

Región Ucayali

1.	Q208	Total de hijos	0.96710
2.	EDUC	Nivel de educación de la mujer	0.93366

Región Lima

1.	EDUC	Nivel de educación de la mujer	0.99199
2.	Q208	Total de hijos	0.98529
3.	QWURBRUR	Área de residencia (urbana o rural)	0.97986
4.	QH21A	Electricidad en el hogar	0.97938

V. RESULTADOS

1. Nivel regional

Cuadro 14

PERÚ: MUJERES UNIDAS, MUJERES ACTUALMENTE USUARIAS Y TASA DE PREVALENCIA, POR REGIONES, 1993

Regiones	Mujeres en edad fértil unidas		Tasa de prevalencia (%)	
	Total	Usuarías	Estimada 1993	ENDES II 1991-1992
Total	3 154 895	1 796 273	56.9	59.0
Loreto	93 597	30 334	32.4	34.8
A.A. Cáceres	280 831	138 486	49.3	52.8
Arequipa	130 920	89 842	68.6	70.6
Chavín	136 010	68 049	50.0	53.8
Grau	218 961	120 921	55.2	56.8
Inka	189 001	77 190	40.8	44.7
J.C. Mariátegui	208 086	114 239	54.9	59.6
Los Libertadores	208 330	79 146	38.0	38.6
Nororiental	350 563	204 760	58.4	61.3
La Libertad	181 355	104 098	57.4	57.9
San Martín	85 704	41 480	48.4	53.7
Ucayali	47 881	24 515	51.2	56.1
Lima-Callao	1 023 656	703 213	68.7	70.9

2. Nivel departamental

Cuadro 15

PERÚ: MUJERES UNIDAS, MUJERES ACTUALMENTE USUARIAS Y TASA DE PREVALENCIA ESTIMADA, SEGÚN DEPARTAMENTOS, 1993

Departamentos	Mujeres en edad fértil unidas		Tasa de prevalencia estimada (%)
	Total	Usuarias	
Total	3 458 336	1 975 475	57.1
Amazonas	350 563	204 760	58.4
Ancash	133 150	66 891	50.2
Apurímac	52 427	17 373	33.1
Arequipa	130 920	89 842	68.6
Ayacucho	64 117	21 023	32.8
Cajamarca	169 671	93 480	55.1
Provincia Const. Callao	95 151	65 844	69.2
Cusco	150 315	62 137	41.3
Huancavelica	53 040	18 040	34.0
Huanuco	92 014	41 536	45.1
Ica	82 380	39 620	48.1
Junín	144 663	75 865	52.4
La Libertad	181 355	104 098	57.4
Lambayeque	133 770	85 360	63.8
Lima	928 505	637 369	68.6
Loreto	93 597	30 334	32.4
Madre de Dios	10 436	4 635	44.4
Moquegua	19 803	12 556	63.4
Pasco	31 630	15 751	49.8
Piura	195 247	107 223	54.9
Puno	153 683	79 410	51.7
San Martín	85 704	41 480	48.4
Tacna	34 600	22 527	65.1
Tumbes	23 714	13 698	57.8
Ucayali	47 881	24 623	51.4

3. Nivel provincial

Cuadro 16
**PERÚ: MUJERES UNIDAS, MUJERES ACTUALMENTE USUARIAS
 Y TASA DE PREVALENCIA, SEGÚN DEPARTAMENTOS
 Y PROVINCIAS, 1993**

Departamentos	Mujeres en edad fértil unidas		Tasa de prevalencia estimada (%)
	Total	Usuarias	
Total	3 458 336	1 975 475	57.1
Departamento Amazonas	350 563	204 760	58.4
Chachapoyas	5 833	3 453	59.2
Bagua	10 345	5 741	55.5
Bongara	2 752	1 591	57.8
Condorcanqui	4 986	2 583	51.8
Luya	5 934	3 169	53.4
Rodríguez de Mendoza	2 822	1 580	56.0
Uctubamba	14 450	7 803	54.0
Departamento Ancash	133 150	66 891	50.2
Huaraz	16 684	8 626	51.7
Aija	995	457	45.9
Antonio Raymondi	2 421	990	40.9
Sunción	1 254	547	43.6
Bolognesi	3 577	1 653	46.2
Carhuaz	5 466	2 503	45.8
Carlos F. Fitzcarrald	2 907	1 174	40.4
Casma	5 205	2 660	51.1
Corongo	1 033	486	47.0
Huari	8 282	3 404	41.1
Huarmey	3 354	1 751	52.2
Huaylas	6 667	3 173	47.6
Mariscal Luzurriaga	3 025	1 231	40.7
Ocos	886	399	45.0
Pallasca	3 309	1 476	44.6
Pomabamba	3 463	1 486	42.9
Recuay	2 300	1 086	47.2
Santa	51 292	29 031	56.6
Sihuas	4 044	1 670	41.3
Yungay	6 986	3 088	44.2

Cuadro 16 (continuación 1)

Departamentos	Mujeres en edad fértil unidas		Tasa de prevalencia estimada (%)
	Total	Usuarías	
Departamento Apurímac	52 427	17 373	33.1
Abancay	13 115	5 613	42.8
Andahuaylas	17 593	5 032	28.6
Antabamba	1 556	538	34.6
Aymaraes	3 759	1 229	32.7
Cotabambas	6 358	1 857	29.2
Chincheros	6 584	1 923	29.2
Graú	3 462	1 181	34.1
Departamento Arequipa	130 920	89 842	68.6
Arequipa	96 394	68 343	70.9
Camana	6 333	4 028	63.6
Caraveli	3 836	2 421	63.1
Castilla	5 165	3 130	60.6
Caylloma	6 602	3 823	57.9
Condesuyos	2 796	1 683	60.2
Islay	7 430	5 097	68.6
La Unión	2 364	1 317	55.7
Departamento Ayacucho	64 117	21 023	32.8
Huamanga	20 905	7 902	37.8
Cangallo	4 317	1 166	27.0
Huanca Sancos	1 325	425	32.1
Huanta	8 713	2 518	28.9
La Mar	10 077	2 620	26.0
Lucanas	6 961	2 541	36.5
Parinacochas	2 823	1 056	37.4
Paucar del Sara Sara	1 107	447	40.4
Sucre	1 597	543	34.0
Víctor Fajardo	3 447	1 003	29.1
Vilcas Huamán	2 845	802	28.2
Departamento Cajamarca	169 671	93 480	55.1
Cajamarca	32 247	18 768	58.2
Cajabamba	8 992	4 982	55.4
Celedín	10 648	5 782	54.3
Chota	21 574	11 650	54.0
Contumaza	4 494	2 494	55.5
Cutervo	18 587	10 000	53.8
Hualgayoc	10 384	5 587	53.8

Cuadro 16 (continuación 2)

Departamentos	Mujeres en edad fértil unidas		Tasa de prevalencia estimada (%)
	Total	Usuarias	
Jaén	23 715	13 328	56.2
San Ignacio	15 372	8 224	53.5
San Marcos	6 805	3 668	53.9
San Miguel	8 063	4 322	53.6
San Pablo	3 255	1 725	53.0
Santa Cruz	5 535	2 950	53.3
Prov. Const. Callao	95 151	65 844	69.2
Departamento Cusco	150 315	62 137	41.3
Cusco	39 752	22 500	56.6
Acomayo	4 231	1 540	36.4
Anta	8 067	3 114	38.6
Calca	8 294	3 011	36.3
Canas	5 426	1 725	31.8
Canchis	13 423	5 718	42.6
Chumbivilcas	9 854	2 828	28.7
Espinar	7 996	3 022	37.8
La Convención	24 011	8 332	34.7
Paruro	5 022	1 557	31.8
Paucartambo	6 419	1 945	30.3
Quispicanchis	11 176	4 001	35.8
Urubamba	6 644	2 844	42.8
Departamento Huancavelica	53 040	18 040	34.0
Huancavelica	14 234	4 711	33.1
Acobamba	6 081	1 618	26.6
Angares	6 121	1 604	26.2
Castrovirreyna	2 322	882	38.0
Churcampa	5 908	1 707	28.9
Huaytara	2 990	1 026	34.3
Tayacaja	15 384	6 492	42.2
Departamento Huanuco	92 014	41 536	45.1
Huanuco	31 018	15 199	49.0
Ambo	7 526	3 138	41.7
Dos de Mayo	14 435	5 889	40.8
Huacaybamba	2 507	975	38.9
Huamalies	7 600	3 093	40.7
Leoncio Prado	15 357	7 863	51.2
Marañón	2 860	1 158	40.5
Pachitea	5 972	2 269	38.0
Puerto Inca	4 739	1 952	41.2

Cuadro 16 (continuación 3)

Departamentos	Mujeres en edad fértil unidas		Tasa de prevalencia estimada (%)
	Total	Usuarías	
Departamento Ica	82 380	39 620	48.1
Ica	34 459	17 677	51.3
Chincha	22 812	10 357	45.4
Nasca	7 891	3 740	47.4
Palpa	1 870	817	43.7
Pisco	15 348	7 029	45.8
Departamento Junín	144 663	75 865	52.4
Huancayo	60 639	34 079	56.2
Concepción	7 959	3 940	49.5
Chanchamayo	17 675	8 184	46.3
Jauja	12 598	6 853	54.4
Junín	5 055	2 492	49.3
Satipo	14 619	6 184	42.3
Tarma	15 760	8 053	51.1
Yauli	10 358	6 080	58.7
Departamento La Libertad	181 355	104 098	57.4
Trujillo	92 758	57 046	61.5
Ascope	15 572	9 250	59.4
Bolívar	2 244	1 135	50.6
Chepén	9 093	5 374	59.1
Julcán	5 059	2 383	47.1
Otuzco	14 791	7 144	48.3
Pacasmayo	12 244	7 334	59.9
Pataz	8 290	3 929	47.4
Sánchez Carrión	14 519	7 056	48.6
Santiago de Chuco	6 785	3 447	50.8
Departamento Lambayeque	133 770	85 360	63.8
Chiclayo	91 382	60 678	66.4
Ferreñafe	13 625	8 230	60.4
Lambayeque	28 763	16 452	57.2
Departamento Lima	928 505	637 369	68.6
Lima	831 155	572 666	68.9
Barranca	16 920	11 218	66.3
Cajatambo	1 122	688	61.3
Canta	1 341	900	67.1
Cañete	21 999	14 761	67.1
Huaral	18 951	12 470	65.8
Huarochari	7 928	5 312	67.0
Huara	23 420	15 691	67.0
Oyon	2 250	1 424	63.3
Yauyos	3 419	2 239	65.5

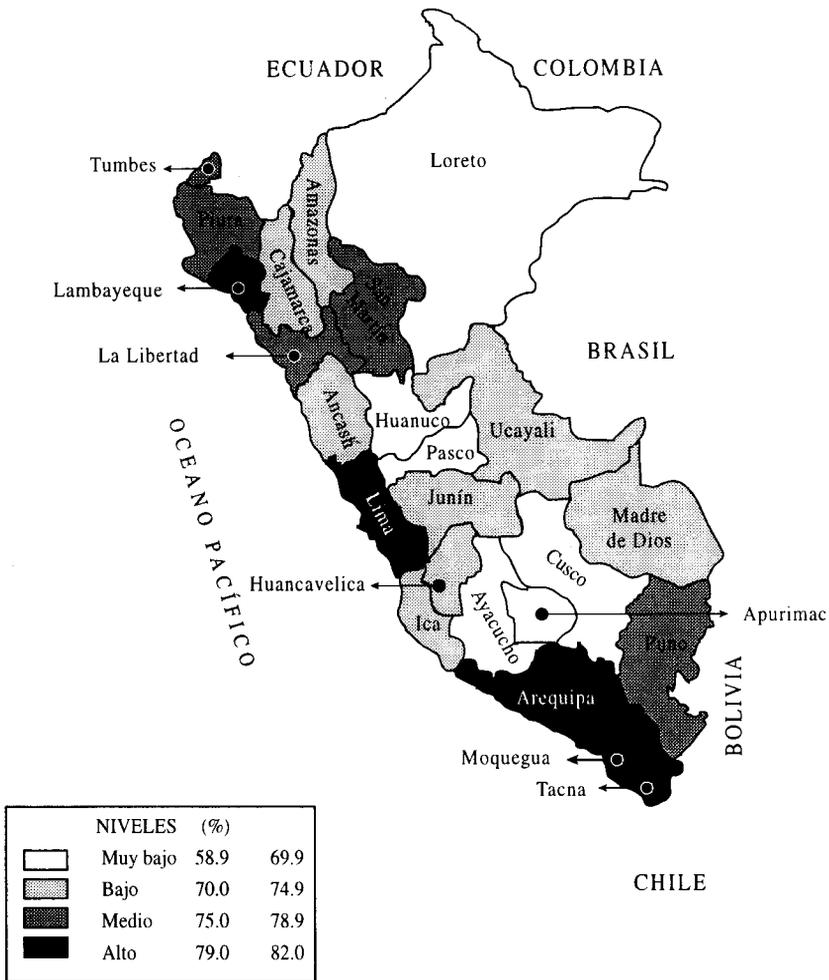
(Cuadro 16 (continuación 4))

Departamentos	Mujeres en edad fértil unidas		Tasa de prevalencia estimada (%)
	Total	Usuarias	
Departamento Loreto	93 597	30 334	32.4
Maynas	54 074	20 764	38.4
Alto Amazonas	15 455	3 879	25.1
Loreto	6 732	1 609	23.9
Mariscal Ramón Castilla	4 342	986	22.7
Requena	6 977	1 688	24.2
Ucayali	6 017	1 408	23.4
Departamento Madre de Dios	10 436	4 635	44.4
Tambopata	7 311	3 429	46.9
Manu	2 185	782	35.8
Tahuamanu	940	424	45.1
Departamento Moquegua	19 803	12 556	63.4
Mariscal Nieto	8 976	5 772	64.3
General Sánchez Cerro	2 497	1 261	50.5
Ilo	8 330	5 523	66.3
Departamento Pasco	31 630	15 751	49.8
Pasco	18 029	9 826	54.5
Daniel Alcides Carrión	4 708	2 057	43.7
Oxapampa	8 893	3 868	43.5
Departamento Piura	195 247	107 223	54.9
Piura	77 840	43 902	56.4
Ayabaca	16 272	7 745	47.6
Huancabamba	14 578	6 852	47.0
Morropón	21 563	11 213	52.0
Paíta	11 478	6 336	55.2
Sullana	34 170	19 374	56.7
Talara	19 346	11 801	61.0
Departamento Puno	153 683	79 410	51.7
Puno	28 978	16 228	56.0
Azangaro	18 574	8 600	46.3
Carabaya	6 509	3 111	47.8
Chucuito	13 194	6 175	46.8
El Collao	11 186	5 291	47.3
Huancane	11 188	5 214	46.6
Lampa	5 732	2 746	47.9

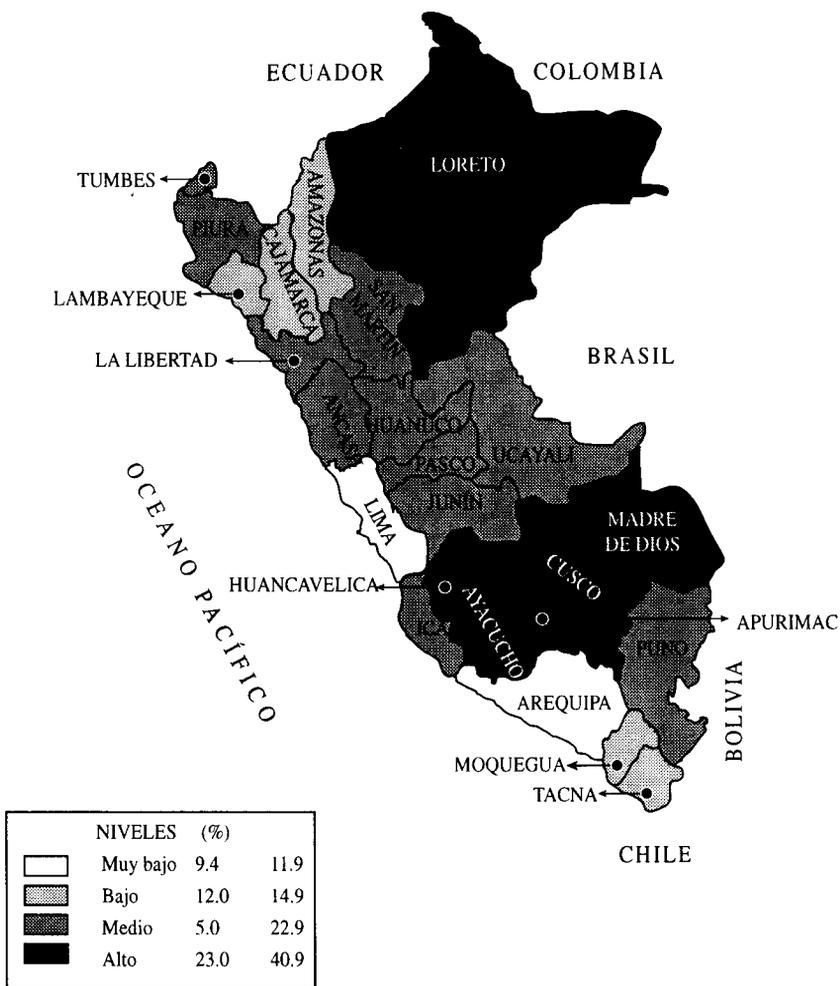
Cuadro 16 (conclusión)

Departamentos	Mujeres en edad fértil unidas		Tasa de prevalencia estimada (%)
	Total	Usuarías	
Melgar	9 678	5 071	52.4
Hoho	4 366	1 987	45.5
San Antonio de Putina	4 199	2 020	48.1
San Román	25 833	16 275	63.0
Sandia	7 254	3 322	45.8
Yunguyo	6 992	3 370	48.2
Departamento San Martín	85 704	41 480	48.4
Moyobamba	10 417	5 021	48.2
Bellavista	5 516	2 659	48.2
El Dorado	3 573	1 676	46.9
Huallaga	3 454	1 641	47.5
Lamas	10 009	4 644	46.4
Mariscal Cáceres	7 624	3 690	48.4
Picota	4 390	2 208	50.3
Rioja	10 672	4 909	46.0
San Martín	18 500	9 639	52.1
Tocache	11 549	5 393	46.7
Departamento Tacna	34 600	22 527	65.1
Tacna	30 266	19 824	65.5
Candarave	1 423	810	56.9
Jorge Basadre	1 922	1 290	67.1
Tarata	989	603	61.0
Departamento Tumbes	23 714	13 698	57.8
Tumbes	17 649	10 307	58.4
Contralmirante Villar	1 976	1 097	55.5
Zarumilla	4 089	2 294	56.1
Departamento Ucayali	47 881	24 623	51.4
Coronel Portillo	37 127	19 863	53.5
Atalaya	4 535	1 946	42.9
Padre Abad	5 788	2 634	45.5
Purus	431	180	41.8

Mapa
**PERÚ: DEMANDA TOTAL DE PLANIFICACIÓN FAMILIAR
 POR DEPARTAMENTOS**



Mapa
**PERÚ: DEMANDA NO SATISFECHA DE PLANIFICACIÓN FAMILIAR
 POR DEPARTAMENTOS**



Validez de los resultados

Con la aplicación del procedimiento descrito se obtuvieron resultados por provincias para las 13 regiones administrativas que considera la ENDES II. Las estimaciones provinciales se expresan en tasas (o proporciones) de mujeres unidas, para cada categoría de la demanda de métodos anticonceptivos, es decir, para la demanda no satisfecha, el uso actual y falla del método. A partir de las proporciones estimadas por el modelo y del número de mujeres unidas (con datos del censo) se calculó la población femenina de cada provincia que demanda planificación familiar. Se dispone así de tasas y del número de mujeres unidas para cada componente de la demanda de métodos anticonceptivos en los planos regional, departamental y provincial.

Errores a nivel regional

La validez de los resultados se verifica mediante la comparación entre las proporciones calculadas con los datos de la ENDES II y las obtenidas en base al modelo. En el caso de la demanda total, tanto en el promedio nacional como en la mayoría de las regiones las cifras son similares, con menos del 2% de diferencia. La región de Ucayali es la que registra la mayor discrepancia (casi 6%), explicada básicamente por las cifras de uso actual; esta región tiene un número reducido de casos encuestados, el que se reduce todavía más por efecto de la ponderación, sugiriendo que los errores de muestreo y de estimación pueden ser importantes.

En el análisis de la demanda no satisfecha, se aprecia que las discrepancias son más homogéneas y de magnitud inferior al 1%. Como ya se ha indicado, las mayores diferencias se presentan en el uso actual, originadas por el número de subclases utilizadas en las estimaciones y por los errores propios del muestreo. En la falla del método, pese al pequeño orden de magnitud y a la forma de cálculo (valor residual), no se registran situaciones llamativas. Un cálculo sencillo para medir las discrepancias entre los valores muestrales y las estimaciones es el error de proyección, con el que obtienen los siguientes valores: demanda total, 0.54; demanda no satisfecha, 0.31, uso actual, 0.79 y falla del método, 0.08.

Comparación con otros indicadores a nivel departamental

Es conveniente observar los resultados estimados de la demanda y sus componentes en el contexto de algunos indicadores demográficos relacionados al uso de anticonceptivos (la tasa global de fecundidad, tasa de analfabetismo y tasa de mortalidad infantil), con la finalidad de resaltar la coherencia en términos globales de los mismos.

Se observa que en los departamentos de mayor tasa de prevalencia de anticonceptivos los niveles de mortalidad infantil son los más bajos, y ese es el caso de Arequipa, Callao, Lima y Lambayeque, entre otros. Inversamente, departamentos que registran bajo uso de anticonceptivos (como Apurímac, Cusco, Huancavelica, Ayacucho y Loreto) registran altas tasas de mortalidad infantil. Una prueba de correlación lineal simple aplicada al total de departamentos da como resultado un alto coeficiente de correlación (-0.82), cifra que muestra una coherencia importante de los resultados. En el caso de la demanda no satisfecha y de la mortalidad infantil, la correlación es directa aunque de menor intensidad (0.76) y se presentan casos como el departamento de Ica, que registra una demanda no satisfecha relativamente alta y una baja mortalidad infantil; existen otros factores, por ejemplo, el bajo porcentaje de analfabetos, que tendría una incidencia mayor en la mortalidad infantil. Relacionando el porcentaje de analfabetos con la prevalencia de anticonceptivos y la demanda no satisfecha se encuentran, como en el caso anterior, relaciones esperadas aunque de menor intensidad. En ambos casos el coeficiente de correlación es de 0.72, con signo negativo para la prevalencia y positivo para la demanda no satisfecha. Estos resultados también muestran la coherencia de las estimaciones en el contexto sociodemográfico. En el análisis de las variables independientes que explicarían la demanda de planificación familiar, no se utilizó el alfabetismo, por considerarse que el nivel educativo tenía mayor relevancia, esta última variable si mostró alto nivel de explicación en la mayoría de las regiones. En todo caso, ambas variables son indicadores del mismo fenómeno educativo.

La mayor correlación (-0.86) se observa entre la fecundidad y el uso de anticonceptivos. La demanda no satisfecha y la fecundidad muestran una relación directa positiva (0.72), que pese a ser menor que la anterior es igualmente importante. Finalmente se tomó el porcentaje de hogares con alguna necesidad básica insatisfecha como un indicador de pobreza a nivel departamental; Como en los casos anteriores, hay una alta correlación inversa entre la prevalencia anticonceptivo y el

indicador de NBI (-0.86). Tomando la demanda no satisfecha y el NBI el coeficiente de correlación es igualmente alto y positivo (0.73).

Comentarios generales

Según estimaciones realizadas con información proveniente de los Censos Nacionales de 1993 y la Encuesta Demográfica y de Salud Familiar (ENDES), en el país existen 5 630 378 mujeres de 15 a 49 años. El 56,1% de este total, es decir, 3 160 000, declararon estar unidas maritalmente (casadas o convivientes). De estas mujeres unidas, el 75,6%, es decir, 2 388 828 estarían demandando servicios de planificación familiar y el 24,4% restante no usa métodos de planificación familiar porque desea un embarazo o son infértiles.

Las mujeres que demandan servicios de planificación familiar, se dividen en 3 grandes grupos: usuarias actuales de algún método, mujeres con demanda no satisfecha y usuarias cuyo método falló. Las mujeres que usan actualmente algún método anticonceptivo, ya sea tradicional o moderno, ascienden a 1 799 290 (56,9%) del total de mujeres unidas. La demanda no satisfecha alcanza a 526 640 mujeres (16,7%) de las mujeres casadas o convivientes. Este grupo de mujeres comprende a las que no desean tener más hijos o que desean postergar su embarazo por lo menos dos años y, a su vez, no usan método anticonceptivo alguno, presumiblemente por falta de información y/o de acceso a los servicios de planificación familiar. Las mujeres que pese a estar usando un método anticonceptivo quedaron embarazadas también forman parte de la demanda y totalizan 62 898 (falla del método).

a) Estimaciones departamentales

El nivel de la demanda total de servicios de planificación familiar es significativo en todos los departamentos del país, llegando al 75,6% el promedio nacional. La demanda total va del 57,8% en Loreto, al 82,0% en Tacna. Son diez los departamentos del país, incluido el Callao, donde esta tasa se encuentra por encima del promedio nacional y quince departamentos los que están por debajo de dicho promedio.

Entre los departamentos que presentan la mayor tasa de demanda total, además de Tacna, se encuentran: Moquegua (81,7%), Callao (80,7%), Lima (79,8%), Arequipa (79,4%), Puno (78,7%), Tumbes (77,1%), Piura (77,0%) y La Libertad (76,3%). La mayoría de estos

departamentos están localizados en la costa del Perú y tiene un mayor desarrollo socioeconómico.

Entre los departamentos que presentan la menor tasa de demanda total, además de Loreto, se encuentran: Huánuco (66,3%), Cusco (68,4%), Pasco y Apurímac (69,7%) y Madre de Dios (70,5%), localizados en la Sierra y Selva de nuestro país y presentan, además, bajos niveles de desarrollo socioeconómico.

De otro lado, también es importante que la demanda total de métodos anticonceptivos sea analizada en valores absolutos, dada la relación directa con la cantidad de insumos necesarios para atender a la población demandante. En un primer grupo, conformado por nueve departamentos, destaca claramente Lima, donde se encuentra la tercera parte de las mujeres unidas de todo el país (741 225 mujeres) que demandan métodos anticonceptivos. Completan el grupo Piura, La Libertad, Cajamarca, Puno, Lambayeque, Arequipa, Junín, Cusco y Ancash, con poblaciones entre cien y doscientas mil mujeres demandantes. El segundo grupo está formado por los otros 16 departamentos, es decir, aquellos con menos de 100 mil mujeres que demandan planificación familiar. Sin embargo, es de resaltar que es en estos departamentos, donde se presentan los mayores niveles de fecundidad y mortalidad infantil del país. Por tal razón, Huánuco, Loreto, Ayacucho, Huancavelica Apurímac y Amazonas, que forman este grupo, deben recibir un trato preferencial en la programación de los servicios de planificación familiar.

A nivel departamental, la tasa de demanda insatisfecha de servicios de planificación familiar es muy diferencial y va de 40.4% en Ayacucho a sólo 9.4% en Arequipa, siendo la tasa promedio nacional de 16.7%. Los departamento con las tasas más altas de demanda insatisfecha son: Ayacucho (40.4%), Huancavelica (36.7%), Apurímac (34.4%), Loreto (25.5%), Madre de Dios (24.6%), Cusco (23.2% y San Martín (21.1%). Como se aprecia de estos siete departamentos, cuatro pertenecen a la Región de la Selva y los otros tres conforman el Trapecio Andino. Asimismo, estos departamentos presentan las más altas tasas de fecundidad y de mortalidad infantil, lo que evidencia su menor desarrollo socioeconómico relativo y en consecuencia el menor acceso a los servicios de salud educación y de la vivienda.

Los departamentos con las menores tasas de demanda insatisfecha son: Arequipa (9.4%), Callao (10.3%), Lima (10.8%), Tacna (12.5%), Moquegua (13.5%) y Lambayeque (14.0%). Todos ellos se ubican en la Región de la Costa y se encuentran entre los que han alcanzado un mayor desarrollo socioeconómico. En estos departamentos, las tasas de

fecundidad y de mortalidad infantil son las más bajas del país y se encuentran por debajo de la tasa promedio nacional.

El análisis de la demanda no satisfecha de planificación familiar, en cifras absolutas, nos muestra que el departamento de Lima, alberga al mayor número de mujeres unidas que necesitan servicios de planificación familiar (100 mil), siguiéndole en volumen Piura (37 mil), Cusco (35 mil), Puno (31 mil), La Libertad (30 mil) y Ancash (27 mil). Estos seis departamentos en conjunto, reúnen a casi la mitad del total de mujeres unidas con demanda insatisfecha.

Los departamentos con menor número de mujeres unidas con demanda insatisfecha son: Madre de Dios (2 600), Moquegua (2 700), Tacna (4 300), Tumbes (4 200) y Pasco (5 400). Estos departamentos, en conjunto, no alcanzan a registrar 20 mil mujeres unidas.

b) Estimaciones a nivel provincial

Así como la demanda de planificación familiar por departamentos nos muestra significativas diferencias interdepartamentales, el análisis a nivel provincial, revela diferencias entre las provincias de un mismo departamento.

El país se caracteriza por tener una composición sociodemográfica heterogénea, con desequilibrios socioeconómicos regionales. En consecuencia, la programación eficiente de las acciones de políticas de población deberá contar con información estadística confiable sobre la población objetivo a niveles desagregados, en este caso, provincial, a fin de focalizar adecuadamente la aplicación de los programas de planificación familiar.

A nivel provincial, la demanda total de servicios de planificación familiar también es muy diferenciada. La tasa más baja corresponde a la provincia Mariscal Ramón Castilla, del departamento de Loreto (50.3%) y la más alta se da en la provincia Jorge Basadre, del Departamento de Tacna (82.6%).

Las provincias con los mayores niveles de demanda son las siguientes: Huaura (Lima) 79.0%, Melgar (Puno) 79.1, Islay (Arequipa) 79.5%, Trujillo (La Libertad) 79.8%, Puno (Puno) 79.9%, Lima (Lima) 80.0%, Arequipa (Arequipa) y Candarave (Tacna) 80.3%, Callao (Callao) 80.7% Chiclayo (Lambayeque) 81.3%, Mariscal Nieto (Moquegua) y San Román (Puno) 82.0%, Tacna (Tacna) 82.1%, Ilo (Moquegua) 82.3% y Jorge Basadre (Tacna) 82.6%. En su conjunto, en estas provincias residen 1 086 000 mil mujeres unidas que demandan servicios de planificación familiar (45.5% de la demanda total). Es

oportuno señalar que la presencia de Lima en este nivel explica el importante volumen de mujeres que demandan esos servicios.

BIBLIOGRAFÍA

- García Núñez, José y otros (1994), "Estimaciones programáticas para áreas geográficas pequeñas", México, D.F., mayo.
- INEI (Instituto Nacional de Estadística e Informática), DHS (Encuestas de Demografía y Salud) (1992), Perú: Encuesta demográfica y de salud familiar, 1991-1992, Lima, septiembre.
- Loza, Gloria y Guillermo Vallenos (1992), "Uso y demanda de métodos anticonceptivos en el Perú", serie OI, N° 90 (LC/DEM/R.196), Santiago de Chile, Encuestas de Demografía y Salud (EDS)/Centro Latinoamericano de Demografía (CELADE).
- Mostajo, Nelly (1981), "Actitudes de la mujer frente a la fecundidad y uso de métodos anticonceptivos", serie D, N° 107, Santiago de Chile, Centro Latinoamericano de Demografía (CELADE).
- Ortiz, J. y E. Alcántara (1988), "Cambios en la fecundidad peruana", Lima, Centro de Investigación en Población Cuzco, Universidad Nacional San Antonio Abad del Cuzco.
- Westoff, C. y L.H. Ochoa (1991), "Unmet need the demand for family planning", Demographic and Health Surveys, Comparative Studies, N° 5, Maryland, Institute for Resource Development/Macro International Columbia.

**COMBINACIÓN DE DATOS CENSALES Y DE ENCUESTAS
PARA ESTUDIAR LAS DIMENSIONES ESPACIALES
DE LA POBREZA: EL CASO DE ECUADOR**

Jesko Hentschel
Banco Mundial

Jean Olson Lanjouw
Universidad de Yaley
y Universidad Libre
de Amsterdam

Peter Lanjouw
Banco Mundial
y Universidad Libre
de Amsterdam

Javier Poggi
Superintendencia
de Banca y Seguros,
Lima, Perú

RESUMEN

Los mapas de pobreza, que proporcionan información sobre la distribución espacial de los niveles de vida, son una herramienta importante para la formulación de políticas y la investigación económica. Estos mapas pueden ser utilizados por los encargados de la formulación de políticas para orientar la asignación de transferencias y servir de base para el diseño de las políticas. También pueden proporcionar los medios necesarios para investigar la relación entre el crecimiento y otras variables económicas, ambientales, sociales, y la distribución en un país determinado. Un obstáculo importante para la elaboración de los mapas de pobreza ha sido que, normalmente, sólo pueden obtenerse datos confiables sobre el ingreso o el consumo mediante encuestas con muestras relativamente pequeñas. Los datos censales tienen la cobertura

requerida, pero generalmente no poseen la información necesaria. En el caso de Ecuador, demostramos cómo pueden combinarse los datos de las encuestas por muestreo con los datos censales para predecir los porcentajes de pobreza de la población cubierta por el censo. Se comprueba que la medición de estos indicadores es bastante precisa, aun para niveles de desagregación relativamente elevados. Sin embargo, si se excede cierto grado de desagregación espacial, los errores estándar aumentan apreciablemente.

Los autores agradecen al Instituto Nacional de Estadística y Censos (INEC) de Ecuador por permitirnos acceder a los microdatos del Censo de 1990. Las discusiones con Uwe Deichmann, Stefan Dercon, Chris Elbers, Francisco Ferreira, Joanna Gomulka, Vassilis Hajivassiliou, Jeffrey Hammer, Berk Ozler y Martin Ravallion nos resultaron muy provechosas. También recibimos comentarios útiles de tres árbitros anónimos, así como de los participantes de un seminario celebrado en los STICERD del London School of Economics, en mayo de 1997, y de un taller denominado "Geographical Targeting for Poverty Reduction and Rural Development" celebrado en el Banco Mundial en noviembre de 1997. Agradecemos también a Jim Shafer por su valiosa ayuda. Las opiniones expresadas en el presente estudio son nuestras y no deben considerarse un reflejo de los puntos de vista del Banco Mundial o de cualesquiera de sus afiliados. Todos los posibles errores son de nuestra responsabilidad.

COMBINING CENSUS AND SURVEY DATA IN STUDYING THE SPATIAL DIMENSIONS OF POVERTY: THE CASE OF ECUADOR

ABSTRACT

Poverty maps, providing information on the spatial distribution of living standards, are an important tool for policy making and economic research. Such maps can be used by policymakers to guide the allocation of transfers and to inform policy design. They can also provide the means to investigate the relationship between growth and other economic, environmental or social outcomes, and distribution within a country. A major impediment to the development of poverty maps has been that the necessary credible income or consumption data are typically available only from relatively small surveys. Census data have the required size, but not generally the required information. We demonstrate for the case of Ecuador how sample survey data can be combined with census data to yield predicted poverty rates over the population covered by the census. These poverty rates are found to be rather precisely measured, even at fairly disaggregated levels. However, beyond a certain level of spatial disaggregation the standard errors rise rapidly.

I. INTRODUCCIÓN

Los mapas de pobreza proporcionan una descripción detallada de la distribución espacial de la pobreza en un país. Pueden tener un valor considerable para los gobiernos, las organizaciones no gubernamentales y las instituciones multilaterales interesadas en fortalecer los impactos del gasto sobre la mitigación de la pobreza. Por ejemplo, pueden utilizarse para orientar la distribución de recursos entre las agencias o los gobiernos locales como primer paso para llegar a los pobres. Hoy día, muchos países en desarrollo utilizan los mapas de pobreza con ese propósito.

Estos mapas también pueden ser una herramienta importante para la investigación. Generalmente, la relación empírica entre la pobreza o la desigualdad y los indicadores del desarrollo, como el crecimiento económico, se examina en el marco de una regresión entre países.¹ Sin embargo, es difícil controlar la enorme heterogeneidad existente entre los países, una heterogeneidad que puede ocultar relaciones genuinas entre las variables de interés. Asimismo, el universo de las experiencias de países que puede utilizarse para comprender los determinantes y los efectos de la distribución del bienestar es limitado. Los estudios realizados en una escala más micro —utilizando una variación de la distribución entre comunidades de un mismo país— ofrecen mejores posibilidades de avanzar en este tema.

Sin embargo, la elaboración de mapas de pobreza se ve obstaculizada por la escasez de datos desagregados. Por ejemplo, si bien muchas veces se prefiere utilizar los indicadores de ingresos o gastos, por lo general no se dispone de la información necesaria para elaborar un mapa muy desagregado de los ingresos o los gastos de un número suficiente de hogares. Mediante las encuestas de medición de los

1 Recientemente, Deininger y Squire (1996) han compilado una gran base de datos internacional con ese fin. Bruno, Ravallion y Squire (1998) la utilizan para analizar la relación entre el crecimiento económico y la desigualdad. Véase también, Alesina y Rodrik (1994) y Fields (1989).

niveles de vida del Banco Mundial que, con distintas variantes, se han aplicado en muchos países en desarrollo, se recopila la información necesaria para realizar mediciones integrales de los ingresos y el consumo, pero sus tamaños de muestra son demasiado pequeños como para permitir un nivel de desagregación que vaya más allá de un simple desglose entre las zonas rurales y urbanas de las grandes divisiones geográficas de un país determinado. Los datos censales no presentan los problemas característicos de las muestras pequeñas, pero por lo general contienen una información más bien limitada. En muchos países latinoamericanos, y también en África y Asia, los mapas de pobreza usados para realizar una clasificación jerárquica de las regiones se han basado en índices de bienestar elaborados realizando algún tipo de combinación de la información censal básica —como el acceso a los servicios públicos, los niveles de la educación, etc.² Estos factores suelen denominarse indicadores de las “necesidades básicas insatisfechas”. Por lo general, se elaboran en forma bastante pragmática y utilizando los exiguos datos cualitativos obtenidos en los censos, sin ajustes en función de la calidad. Como señalamos más adelante, estos indicadores suelen ser sustitutos inadecuados de los niveles de consumo de los hogares. Usando datos detallados de las encuestas de hogares de Ecuador, demostramos que un indicador rudimentario de las necesidades básicas y una medida integral del consumo generan clasificaciones jerárquicas bien distintas del bienestar de los hogares. Procedemos a analizar en qué medida pueden mejorarse los mapas elaborados con datos censales cuando se desea utilizar indicadores de bienestar según ingresos o consumo.

En ciertas situaciones, quizás interese ir más allá de la noción de bienestar reflejada solamente por las posibilidades de acceso a los recursos y examinar también, en forma explícita, otros componentes del bienestar. Si, por ejemplo, el interés se centra en los programas de educación, quizás se desee reconocer el valor intrínseco de la educación, sin limitar el análisis al papel que pueda desempeñar en el mejoramiento de los niveles de ingresos o de consumo. En este caso, en el indicador pertinente de bienestar puede otorgarse a la educación una ponderación mayor que la directamente asociada al consumo o a los ingresos. Si por otra parte, por ejemplo, se desea hacer hincapié en la forma de compensar los hogares por un cambio general en los niveles

2 Para una descripción reciente de la derivación de estos mapas véanse: Banco Mundial (1996) para Ecuador, Gobierno de El Salvador (1995) para El Salvador y FONCODES (1995) para Perú. Entre los países latinoamericanos que usan ese tipo de mapas para orientar la asignación del gasto público destinado al sector social, también cabe mencionar a Colombia, Honduras y Venezuela.

de precios, podría preferirse una medida del bienestar basada en un índice mucho más focalizado en el consumo. En general, se puede elaborar un indicador de las necesidades básicas asignando ponderaciones que reflejen la relación de cada una de las variables disponibles con los recursos totales del hogar, así como cualquier contribución directa al bienestar no comprendida en los ingresos. Si se procede de esta manera, no sería dable esperar que las clasificaciones jerárquicas obtenidas en función de ese indicador correspondan a los indicadores de consumo. Sin embargo, no cabe duda de que sería deseable que las mediciones más amplias también reflejen, en cierta medida, los niveles reales de consumo. En consecuencia, los resultados de este estudio, que demuestran cómo se puede lograr este objetivo de manera más confiable en un nivel desagregado, son útiles aun cuando se busque obtener una medida más amplia del bienestar. Sólo sería necesario combinar de algún modo estas predicciones más precisas de la pobreza —basadas en el consumo— con otros indicadores considerados pertinentes para el tema de política que se examina.

Nuestro enfoque es el siguiente: estimamos modelos del gasto de consumo usando los datos de las encuestas de niveles de vida de Ecuador, restringiendo el conjunto de variables explicativas a las que también pueden obtenerse del censo más reciente de ese país. Luego aplicamos las estimaciones de los parámetros de estos modelos a los datos censales para predecir la probabilidad de que un hogar censado esté en situación de pobreza. Verificamos la validez de nuestro enfoque estimando la incidencia de la pobreza en seis grandes regiones y comparándolas con las tasas estimadas utilizando solamente la encuesta de hogares. Se observa una coincidencia estrecha entre las tasas de pobreza obtenidas a partir de cada uno de estos conjuntos de datos.

Analizamos algunos de los problemas estadísticos que surgen del hecho de que estamos trabajando con predicciones de porcentajes de pobreza. Mediante el enfoque descrito, se obtienen estimaciones insesgadas de la incidencia de la pobreza, de modo que, en promedio, los errores de predicción son nulos. Sin embargo, las estimaciones de pobreza tienen un cierto error estándar y es importante calcularlo al mismo tiempo que las tasas de pobreza. Demostramos que, afortunadamente, en nuestro ejemplo sobre el Ecuador, los errores estándar son pequeños para los niveles de desagregación territorial que pueden llegar a tener importancia práctica. Sin embargo, también demostramos que estos errores llegan a ser bastante importantes cuando se calculan los porcentajes de pobreza para grupos muy pequeños de

hogares. Esto sirve para advertirnos sobre los inconvenientes de excederse en la desagregación espacial de los mapas de pobreza.³

En la sección siguiente describimos los datos de Ecuador utilizados para ilustrar este análisis. Luego examinamos las diferencias de focalización implícitas en una asignación basada en un indicador de las necesidades básicas y en una asignación que tome en cuenta el gasto de consumo. En la sección III estimamos modelos de gasto y después predecimos la probabilidad de que cada uno de los hogares censados esté en situación de pobreza. Sobre esta base, estimamos luego las tasas agregadas de pobreza y las comparamos con las tasas obtenidas mediante la encuesta de hogares. En la sección IV desarrollamos mapas de pobreza simples de las provincias de Ecuador y mostramos que la pobreza en ese país varía considerablemente de una provincia a otra y entre las zonas rurales y urbanas. También demostramos que, a medida que se aumenta el nivel de desagregación de los mapas de pobreza, el intervalo de confianza de las tasas de pobreza calculadas se amplía. En la sección V se formulan observaciones finales y sugerencias para futuras investigaciones.

II. FOCALIZACIÓN DE LA POBREZA SOBRE LA BASE DE UN INDICADOR DE LAS NECESIDADES BÁSICAS O DE UN INDICADOR SEGÚN CONSUMO EN ECUADOR

En esta sección examinamos la eficacia de un indicador de las necesidades básicas —en función de su capacidad para identificar a los pobres cuando el indicador de la pobreza es el gasto de consumo.⁴ El indicador de las necesidades básicas que estamos analizando fue

3 Por ejemplo, sería bastante incorrecto utilizar nuestra metodología para identificar *hogares* pobres en forma individual.

4 El consumo es un indicador imperfecto del nivel de vida, pero una medición integral del gasto permite captar con bastante precisión el bienestar *logrado* por un hogar, conforme a su propia definición del conjunto de bienes y servicios necesarios. La elección entre el ingreso y el consumo también es un tema que merece ser analizado. En los países en desarrollo, el argumento más convincente en favor del consumo probablemente sea que, generalmente, resulta más fácil medirlo con precisión. El hecho de que el consumo no sufra variaciones estacionales importantes, e incluso se mantenga aproximadamente constante de un año a otro, significa que probablemente resulte un indicador más adecuado del nivel de vida a largo plazo que una medición del ingreso actual (aunque con respecto a este tema puede ser conveniente consultar a Chaudhuri y Ravallion, 1994). Para otros análisis, véanse Atkinson (1989), Ravallion (1994) y Sen (1984). Hentschel y Lanjouw (1996) y Lanjouw y Lanjouw (1997) también profundizan el análisis de las ventajas de un indicador *integral* del gasto de consumo como indicador de bienestar.

desarrollado en 1994 por el Instituto Nacional de Estadística de Ecuador (INEC) en respuesta a un pedido concreto del Gobierno para que desarrollara un directorio de hogares pobres. Este directorio debía ser utilizado para focalizar las transferencias compensatorias a los hogares pobres para el caso de que se produjera un aumento del precio de la gasolina si el Gobierno eliminaba el subsidio a este combustible. En realidad, este programa nunca llegó a ejecutarse y no queremos transmitir la impresión de que el INEC consideraba que este indicador de las necesidades básicas fuera otra cosa que una medida bastante rudimentaria, desarrollada para satisfacer un pedido urgente del Gobierno formulado con muy poco preaviso. Sin embargo, el enfoque adoptado por el INEC en la elaboración del indicador de las necesidades básicas se asemeja al que se ha utilizado en muchos países y en consecuencia constituye un ejemplo útil.

El indicador de las necesidades básicas del INEC fue elaborado a nivel de hogares y consiste en una combinación ponderada de cinco variables que comprendían el acceso al agua y a los servicios del saneamiento y de eliminación de residuos, la educación (del jefe del hogar) y un índice de hacinamiento (el número de personas por dormitorio).⁵ Se asignó a cada uno de los servicios un número de puntos determinado en función de su disponibilidad y su tipo o nivel. La asignación de puntos a cada servicio se realizó sobre la base de criterios subjetivos y no como resultado de un análisis empírico. En el cuadro 1 se muestra el sistema de ponderación utilizado.

El valor del indicador de las necesidades básicas de cada hogar se calculó, sencillamente, como la suma de los puntos correspondientes a cada servicio. Se consideró que el hogar era tanto más pobre cuanto menor fuera su puntaje.

Usando los datos de una encuesta de hogares reciente, podemos examinar la eficacia del indicador de las necesidades básicas para identificar a los pobres en relación con el consumo. La Encuesta de Condiciones de Vida (ECV) realizada en Ecuador en 1994 es una encuesta de hogares representativa a nivel nacional y elaborada siguiendo de cerca el modelo de las encuestas de medición de los niveles de vida del Banco Mundial. Proporciona información detallada de cada hogar, sobre una amplia gama de temas, incluso el consumo de alimentos, de bienes y artículos no alimenticios, las actividades

5 En otros trabajos sobre mapas de pobreza para el Ecuador, el INEC ha experimentado con una gama de variables más amplia. En El Salvador, el Gobierno ha elaborado un mapa de la pobreza utilizando 12 variables (Gobierno de El Salvador, 1995).

Cuadro 1
PUNTOS ASIGNADOS A CADA SERVICIO INCLUIDO EN EL INDICADOR
DE NECESIDADES BÁSICAS DEL INEC

Servicio y nivel	Suministro de agua	Saneamiento	Eliminación de residuos	Educación	Hacinamiento
1	100	100	100	100	100
2	50	50	50	50	75
3	25	25	25	25	50
4	0	0	0	0	25
5	-	-	-	-	0

Clave:

Agua:

Saneamiento:

Eliminación de residuos:

Educación (del jefe del hogar):

Hacinamiento (personas por dormitorio):

1 = red pública; 2 = camión de agua; 3 = pozo; 4 = otros.

1 = letrina de cisterna y sifón en la casa; 2 = letrina sin cisterna y sifón en la casa; 3 = compartido; 4 = otros.

1 = camión recolector; 2 = residuos quemados o enterrados; 3 = descarte de los residuos; 4 = otros

1 = terciaria; 2 = secundaria; 3 = primaria o alfabetizado; 4 = otros o no se conoce.

1 = uno o menos; 2 = uno a dos; 3 = dos a tres; 4 = tres a cuatro; 5 = más de cuatro.

laborales, el acceso a servicios como la educación y la salud, las prácticas agrícolas y las actividades empresariales del hogar. La encuesta fue realizada en Ecuador por el Servicio Ecuatoriano de Capacitación (SECAP) durante el período comprendido entre junio y septiembre de 1994. En total, se encuestaron más de 4 500 hogares y, tras realizar verificaciones de depuración y consistencia de los datos, se pudo efectuar el análisis con información sobre 4 391 hogares.⁶ El conjunto de datos de la ECV ha sido analizado por el Banco Mundial en un estudio detallado de la pobreza en Ecuador (Banco Mundial, 1996). Hentschel y Lanjouw (1996) calcularon el monto total del consumo de cada hogar y todas las comparaciones sobre el bienestar de

6 Al diseño de la encuesta se incorporaron métodos de clasificación y estratificación que tuvieran en cuenta las tres zonas agroclimáticas principales del país, con un desglose entre zonas rurales y urbanas. También se incluyó un sobremuestreo de las dos ciudades más importantes de Ecuador: Quito y Guayaquil. Se encuestaron 1 374 hogares rurales. Se incorporaron a los datos factores de expansión de los hogares a fin de poder hacer inferencias sobre agregados de población.

los hogares del Informe sobre la Pobreza en Ecuador del Banco Mundial (Banco Mundial, 1996) se basaron en ese criterio.⁷

En el cuadro 2 comparamos la pobreza por región y por zona, usando los indicadores de las necesidades básicas y del consumo. Como no se desarrolló una línea de la pobreza específica para el indicador de las necesidades básicas, las tasas de pobreza deben obtenerse por inferencia. Esto se logra igualando la incidencia nacional de la pobreza obtenida mediante este indicador con el índice de recuento calculado usando el consumo per cápita y la línea de la pobreza en relación con el consumo —de 45 476 sucres por persona y por quincena (aproximadamente 1.50 dólares de los Estados Unidos por persona por día)— desarrollada por el Banco Mundial (1996).

En consecuencia, nos preguntamos cómo difiere la clasificación jerárquica regional de la pobreza cuando se la define usando estos dos indicadores, pero manteniendo constante la fracción total de la población identificada como pobre. Sólo establecemos una distinción entre las zonas rurales y urbanas y las tres principales regiones agroclimáticas del país.

Con este nivel de agregación, las clasificaciones jerárquicas derivadas de las dos definiciones posibles de bienestar son las mismas, pero las diferencias regionales son mucho más acentuadas cuando se aplica el indicador de las necesidades básicas. Utilizando este último indicador, las zonas rurales parecen más pobres que aplicando el indicador del consumo, mientras que las zonas urbanas parecen menos pobres. Si consideramos las zonas rurales, las de las regiones de Oriente y la Costa son más pobres que las de la Sierra, mientras que en las zonas urbanas, la región de la Costa es la más pobre, seguida por la región de la Sierra y luego por la de Oriente. Como se ha destacado en el documento del Banco Mundial (1996), si se utiliza el criterio del consumo, las clasificaciones en las zonas rurales y urbanas de las regiones de la Costa y de la Sierra son sumamente inestables y pueden invertirse fácilmente según donde se trace la línea de la pobreza y según se elija realizar el análisis con una medida alternativa de la pobreza que no sea el índice de incidencia. Con el criterio de las necesidades básicas, se tiene la impresión de que las diferencias de bienestar entre las regiones son inequívocas.

7 El gasto también se ajustó a fin de tener en cuenta las variaciones regionales del costo de vida sobre la base de un índice de precios de los alimentos de Laspeyres que refleja los patrones de consumo de los pobres.

Cuadro 2
INCIDENCIA DE LA POBREZA SEGÚN DISTINTAS DEFINICIONES
DE BIENESTAR

		Consumo per cápita	Indicador de "necesidades básicas"
Costa			
	Urbana	0.26	0.18
	Rural	0.49	0.76
	Total	0.35	0.39
Sierra			
	Urbana	0.22	0.04
	Rural	0.43	0.50
	Total	0.33	0.28
Oriente			
	Urbana	0.20	0.03
	Rural	0.67	0.76
	Total	0.59	0.65
Nacional			
	Urbana	0.25	0.13
	Rural	0.47	0.62
	Total	0.35	0.35

Nota: Los cálculos se basan en dos indicadores alternativos de bienestar aplicados a los datos de las ECV de los hogares.

Finalmente, se hizo un análisis que trascendiera las comparaciones regionales y un cotejo de la eficacia de los dos indicadores a nivel de hogares. A tal fin, aplicamos el diseño utilizado para la intervención programada, considerando que los beneficiarios previstos del programa integraban el quintil inferior de los hogares. Realizamos el experimento siguiente: calculamos el número total de hogares representados en los datos de la ECV y determinamos que la quinta parte del total estaba constituida por cerca de 450 000 hogares. A continuación, calculamos el número total de puntos correspondientes a caltmo, calculamos el porcentaje de hogares beneficiarios correspondiente a cada quintil del gasto per cápita de los hogares. Como el conjunto de beneficiarios

previsto es el primer quintil, el porcentaje correspondiente a este grupo indicará la eficiencia del indicador de las necesidades básicas para identificar a este grupo específico de beneficiarios. Asimismo, si todos los hogares recibieran la misma cantidad de dinero, el porcentaje de beneficiarios también representa el porcentaje de los recursos que llegaría al grupo focalizado. Los resultados de este experimento se indican en el cuadro 3.

Cuadro 3
**DISTRIBUCIÓN DEL QUINTIL INFERIOR, CONFORME A LOS CRITERIOS
 DE LAS NECESIDADES BÁSICAS, SEGÚN LOS QUINTILES
 DEL GASTO DE CONSUMO**

Quintil en función del consumo real per cápita	Porcentaje de hogares beneficiarios sobre la base del indicador de necesidades básicas	
	Porcentaje	Valor acumulado
20% más pobre	41.4	41.4
20% a 40%	29.5	70.9
40% a 60%	19.5	91.4
60% a 80%	8.0	98.4
20% más rico	1.6	100.0

En el cuadro 3 podemos ver que sólo el 41.4% de los hogares identificados con el criterio de las necesidades básicas como integrantes del quintil inferior de todos los hogares corresponden, de hecho, a ese quintil, utilizando un criterio de consumo. De este modo, la fuga de recursos de una asignación basada en este criterio sería muy elevada—casi el 60% de los recursos se distribuiría entre beneficiarios no programados y 10% se distribuiría a los dos quintiles superiores.⁸

8 La eficacia del indicador de necesidades básicas no sería tan baja si el criterio de asignación se orientara, digamos, sólo a las zonas urbanas o se aplicara otro límite a la distribución.

III. PREDICCIÓN DE LA POBREZA

Para dotar de una base más analítica al esquema de ponderación utilizado para evaluar la pobreza, consideramos aquí la posibilidad de imputar los niveles de consumo de los hogares usando datos censales para constituir la base de un mapa de pobreza.⁹ Sólo puede utilizarse este procedimiento si los datos cumplen con determinados requisitos. En primer lugar, es preciso disponer de una encuesta de hogares, como la ECV de Ecuador, y ésta debe corresponder, aproximadamente, al mismo período abarcado por el censo. En segundo lugar, el análisis requiere contar con datos censales a nivel individual.

Para realizar este análisis, se nos ha permitido acceder a los datos censales de 1990 de Ecuador, que abarcan aproximadamente 2 millones de hogares. Aunque los datos de la ECV de 1994 se recopilaban cuatro años después del censo, en este período el crecimiento en Ecuador fue relativamente lento y la tasa de inflación baja, de modo que es razonable suponer que los patrones de consumo se mantuvieron constantes.

La finalidad subyacente del método propuesto en este trabajo es similar a la de los procedimientos sintéticos de estimación para áreas pequeñas aplicados a variables demográficas y en las economías geográficamente desagregadas.¹⁰ En estos casos, el interés radica en la derivación de los atributos no observados de las áreas, como el promedio o valores totales, expresados usualmente en forma de proporciones (Farrell y otros, 1997). Por ejemplo, si se conocen las variaciones de población en un área grande, las técnicas de estimación de áreas pequeñas permiten calcular variaciones de población en niveles geográficos más bajos basándose en la postulación de determinadas relaciones funcionales. Una diferencia importante del método propuesto

9 Si bien la imputación de observaciones faltantes en una muestra es un procedimiento bastante habitual (véase, por ejemplo, Paulin y Ferraro, 1994), la imputación fuera de muestra que combina distintos conjuntos de datos es menos frecuente. En un artículo reciente, Bramley y Smart (1996) combinan información de la Encuesta de Gastos Familiares en Gran Bretaña con información censal para estimar las distribuciones de los ingresos a nivel local. Sin embargo, estos autores no tuvieron acceso a datos de nivel unitario de ninguna de las dos fuentes. No derivaron las distribuciones de ingresos locales a partir de los ingresos predichos de los hogares sino de estimaciones de los promedios de ingresos y de las características de las distribuciones.

10 Para un resumen de este tema véase Purcell y Kish (1980) y para una aplicación de las estimaciones de áreas pequeñas para las estadísticas económicas véase Isaki (1990).

en este trabajo es que predecimos la variable de interés (consumo) a nivel unitario (el hogar) y basamos las estadísticas agregadas en estas predicciones.¹¹

Estimación de modelos de consumo

Para imputar gastos utilizando datos censales, el primer paso es estimar un modelo de consumo usando datos de la encuesta de hogares. Evidentemente, las únicas variables que se pueden utilizar para predecir el consumo son las que también están disponibles en el censo. En el caso de Ecuador, este conjunto de variables predictivas potenciales comprende algunas variables demográficas, como el tamaño del hogar y su composición en términos de edad y de género; la información sobre la educación y la ocupación de cada miembro de la familia; los datos sobre la calidad de la vivienda (materiales de construcción, tamaño); el acceso a los servicios públicos, como la electricidad y el agua; el principal idioma hablado en el hogar; y la ubicación de la vivienda. (Véase el cuadro 1 del apéndice sobre estadísticas comparativas resumidas de los dos conjuntos de datos.) Después de definir un conjunto de variables *dummy*, los términos de interacción y los términos de orden superior, se dispuso un total de 48 variables predictivas para las regresiones.

Se estimaron distintos modelos para cada región (Costa, Sierra y Oriente) y, en cada una de ellas, se distinguió entre las zonas urbanas y rurales. Para Guayaquil y Quito se realizaron estimaciones por separado ya que en la ECV se había realizado un sobremuestreo de estas dos ciudades.¹² La variable dependiente en una de las regresiones fue el logaritmo del gasto de consumo per cápita para el hogar i , $\ln y_i$:

$$\ln y_i = X_i' \beta + \varepsilon_i \quad \varepsilon_i \sim N(0, \sigma^2) \quad [1]$$

siendo el vector de variables independientes X_i común a la ECV y al censo y ε_i es un término de perturbación aleatorio. Los modelos se estimaron utilizando mínimos cuadrados ponderados, considerando

11 Últimamente, la combinación de información proveniente de distintos conjuntos de datos ha suscitado interés en la bibliografía (por ejemplo, Arellano y Meghir, 1992; Angrist y Krueger, 1992; Lusardi, 1996; e Imbens y Hellrestein, 1998). Por lo general, estos estudios combinan varias encuestas de hogares en lugar de combinar datos censales y de encuestas y no se han utilizado para abordar el tema de las estimaciones espaciales de la pobreza.

como ponderadores los correspondientes a las muestras de hogares encuestados como ponderaciones. El poder explicativo de los ocho modelos osciló entre un valor de $R^2 = 0.46$ para las zonas rurales de la Sierra y $R^2 = 0.74$ para la zona rural de Oriente. Los R^2 para los modelos urbanos variaron entre 0.55 (Quito) y 0.64 (zonas urbanas de la Sierra).¹³

-
- 12 Resulta interesante examinar de qué manera se relacionan las ponderaciones de las necesidades básicas indicadas en el cuadro 1 con las que están implícitas en las estimaciones de los coeficientes de regresión. Quito es un ejemplo característico. En la clasificación de necesidades básicas, el aumento del bienestar correspondiente a una disminución del número de personas por dormitorio de cuatro a tres se considera *equivalente* al que correspondería a un aumento en la educación del jefe del hogar del nivel primario al secundario. En la regresión del consumo, las estimaciones permiten inferir que un aumento en la educación del jefe del hogar del nivel primario al secundario está asociado con un incremento en el consumo de 30%, mientras que una disminución del número de personas por dormitorio de cuatro a tres corresponde sólo a un aumento del consumo de 6.7%. Un aumento en la educación del nivel secundario al terciario también corresponde a un incremento del consumo del 30%. Sin embargo, en este caso una disminución del número de personas por dormitorio de tres a uno que, conforme a las ponderaciones de las necesidades básicas tienen un efecto sobre el bienestar equivalente al de la educación universitaria, está asociado a un aumento del consumo de 47.6%, mucho mayor que los anteriores. Este patrón se observa en todas las regiones. De esto puede deducirse que cuanto mayores sean los niveles de hacinamiento y menores los niveles de educación, en el sistema de las necesidades básicas se asigna una ponderación mayor a una reducción del hacinamiento que a un aumento en el nivel de educación, si se la compara con la que correspondería a la regresión del consumo. Si, como se explica en la introducción, las ponderaciones de las necesidades básicas tienen por objeto reflejar las relaciones de ambas variables con el consumo global, así como un ajuste con respecto a su valor intrínseco, el sistema de ponderaciones implícito en la clasificación de las necesidades básicas parece indicar un juicio de valor según el cual el hecho de que el jefe del hogar esté alfabetizado o tenga estudios primarios es menos importante que reducir el hacinamiento en los dormitorios.
- 13 Por motivos de espacio, no se indica aquí el conjunto completo de estimaciones de los parámetros, de los errores estándar y de los diagnósticos para los ocho modelos de regresión, pero pueden solicitarse a los autores. La especificación correcta de la forma funcional precisa del término de perturbación de la regresión de consumo es importante para calcular la incidencia de pobreza de segunda etapa. De este modo, verificamos la hipótesis de normalidad formulada en la ecuación [1]. En tres de las ocho regiones no podemos rechazar la normalidad de los términos de perturbación utilizando la prueba de Shapiro-Wilk y las pruebas combinadas de asimetría y curtosis (todos los valores de $p > 0.15$). Un análisis más minucioso de los valores residuales demostró que, en las otras regiones, la falta de normalidad se debía a unos pocos valores atípicos en ambas colas de la distribución. Es muy posible que esto se deba a errores de medición. (Por ejemplo, en un caso, el consumo máximo era seis veces mayor que el inmediatamente siguiente.) Tras ajustar estas pocas observaciones —sólo 13 de un total de 4 635— no fue posible rechazar la hipótesis de normalidad a los niveles de significación convencionales en ninguna de las regiones. Estas pequeñas desviaciones de la hipótesis de normalidad del término de perturbación no deberían tener efectos significativos en la exactitud de los resultados que siguen. Además, salvo en el caso de Guayaquil, no pudimos rechazar (para $\alpha = 0.10$) la hipótesis nula de homocedasticidad con respecto a la alternativa de heterocedasticidad respecto del conjunto integral de variables independientes.

Antes de pasar a la segunda etapa, que significa aplicar los modelos a los datos censales, verificamos si las predicciones del consumo —realizadas sobre la base de la encuesta— mejorarían la focalización con respecto al indicador de necesidades básicas examinado anteriormente. Si bien, como ya se ha señalado, obtuvimos ajustes bastante razonables para las regresiones transversales, los coeficientes de determinación fueron significativamente menores que uno. Para evaluar la eficacia del modelo, realizamos un experimento similar al indicado en el cuadro 3, en el que el indicador de necesidades básicas se comparó con el dato de consumo. En el cuadro 4 se indican los resultados obtenidos comparando los niveles de consumo con las predicciones.

En ese cuadro podemos ver que los modelos de predicción formulados sobre la base del consumo son, de hecho, más eficaces para identificar los hogares más pobres que el indicador de las necesidades básicas. La primera prueba consistió en utilizar en los modelos de predicción la muestra completa de hogares aplicando a ésta las estimaciones de los parámetros. En el cuadro 4 se observa que la eficacia de la focalización mejora cerca de un 59.9% para el quintil inferior conforme a las predicciones de consumo y lo mismo ocurre en ese quintil si se considera el consumo observado. La segunda prueba fue bastante más exigente. En ella, dividimos la encuesta de hogares por la mitad —en forma aleatoria—, estimamos el modelo de consumo utilizando sólo la mitad de los datos de la encuesta y realizamos una predicción del consumo en la otra mitad (predicción fuera de muestra). Como cabía esperar, en esta prueba el mejoramiento con respecto al indicador de necesidades básicas es mucho menos notable. Sin embargo, si el objetivo es concentrarse en el quintil inferior de la población, este enfoque también mejora la eficacia de la focalización del 41.4% (según las necesidades básicas) al 51.0% según la estimación del consumo.

Cuadro 4
**DISTRIBUCIÓN DEL QUINTIL INFERIOR UTILIZANDO
 LAS PREDICCIONES DE CONSUMO EN TODOS
 LOS QUINTILES DEL CONSUMO
 OBSERVADO**

Quintil de consumo per cápita	Porcentaje de hogares beneficiarios (sobre la base de las predicciones de consumo)		Indicador de necesidades básicas (del cuadro 3)
	Dentro de la muestra ^a	Fuera de la muestra ^b	
20% más pobre	59.9	51.0	41.4
20% a 40%	22.0	27.0	29.5
40% a 60%	13.8	13.1	19.5
60% a 80%	3.9	8.0	8.0
20% más rico	0.2	0.9	1.6

^a En el experimento dentro de muestra se derivaron predicciones de consumo por hogares a partir de modelos estimados utilizando la muestra de hogares completa, aplicando nuevamente las estimaciones de los parámetros de la muestra completa.

^b El experimento realizado fuera de muestra consistió en la estimación de modelos para una submuestra del estudio de medición de los niveles de vida, utilizando luego las estimaciones de los parámetros resultantes para predecir el consumo para el resto de la muestra.

Predicciones de pobreza

A continuación procedemos al segundo paso del experimento de imputación, que consiste en aplicar las estimaciones de los parámetros de las regresiones —la totalidad de la muestra de hogares— a los datos del censo. Para cada hogar censado, se multiplicaron las estimaciones de los parámetros de la regresión pertinente —determinadas por la ubicación del hogar— por las características de aquel, a fin de obtener un valor imputado (del logaritmo) del gasto de consumo per cápita. A continuación estimamos la probabilidad de que el hogar fuera pobre teniendo en cuenta el hecho de que el consumo no podía explicarse totalmente mediante el modelo (los R^2 nunca eran igual a 1) y que las

predicciones de consumo se basaban en los datos de la muestra. Por último, la incidencia de la pobreza se calculó como el promedio de las estimaciones correspondientes a los hogares para la población de una región determinada del censo.¹⁴

Más rigurosamente, dada una línea de pobreza z , el indicador de pobreza P_i de un hogar i es:

$$P_i = 1 \text{ si } \ln y_i < \ln z; \text{ de lo contrario } P_i = 0 \quad [2]$$

Utilizando el modelo de consumo de la ecuación [1], la pobreza esperada de un hogar i con las características observables X_i es:

$$E[P_i | X_i, \beta, \delta] = \Phi \left[\frac{\ln z - X_i' \beta}{\delta} \right] \quad [3]$$

donde Φ es una distribución normal estándar acumulativa. Habida cuenta de que estamos trabajando con el indicador de incidencia de la pobreza [2], el valor de la ecuación [3] es, sencillamente, la probabilidad de que un hogar con las características observables X_i sea pobre.¹⁵ Estimamos el valor de la ecuación [1] para obtener estimaciones de $\hat{\beta}$, el vector de coeficientes, y de σ . De este modo, el estimador de la pobreza esperada del hogar i del censo es:

$$P_i^* = \hat{E}[P_i | X_i, \hat{\beta}, \delta] = \Phi \left[\frac{\ln z - X_i' \hat{\beta}}{\delta} \right] \quad [4]$$

14 Nuestro análisis se realizará sobre la base de una única medida de pobreza —la incidencia— y una única línea de pobreza. Sin embargo, sería posible realizar una clasificación jerárquica de las regiones teniendo en cuenta otras medidas de la pobreza o de la desigualdad y experimentar con varias líneas de pobreza (véase Elbers, Lanjouw y Lanjouw, 1999). El análisis también se realizará teniendo en cuenta la incidencia de la pobreza de los hogares. Para calcular la incidencia a nivel individual es preciso ponderar cada observación a nivel de hogar según el tamaño de éste (véase la nota 17). Las cifras consignadas indicadas en los cuadros corresponden a esos totales ponderados.

15 Es decir, si se extrajeran infinitas muestras de una población de hogares, la tasa de pobreza resultante para los hogares con características observables X_i estaría dada por la ecuación [3]. Cabe señalar que, en general, este valor no será igual a la tasa de pobreza real, que es una muestra de esta población infinita y depende del valor de ε_i .

que, por ser una función continua de estimadores consistentes es, en sí mismo, un estimador consistente de $E[P_i]$.

La pobreza regional P es

$$P = \frac{1}{N} \sum_{i=1}^N P_i, \quad [5]$$

donde N es el número de hogares de la región y la pobreza esperada es

$$E[P | X, \beta, \sigma] = \frac{1}{N} \sum_{i=1}^N E[P_i | X_i, \beta, \sigma]. \quad [6]$$

Por lo tanto, la incidencia de la pobreza predicha, P^* , teniendo en cuenta el modelo de consumo estimado, es

$$P^* = \hat{E}[P | X, \hat{\beta}, \hat{\sigma}] = \frac{1}{N} \sum_{i=1}^N \hat{O} \left[\frac{\ln z - X_i' \hat{\beta}}{\hat{\sigma}} \right]. \quad [7]$$

Obsérvese que calculamos la incidencia de la pobreza como el promedio de las probabilidades de que un hogar sea pobre en lugar de limitarnos a contar los hogares cuyo gasto predicho está por debajo de la línea de la pobreza, z . Este último enfoque produciría estimaciones sesgadas de las tasas de pobreza (véase más adelante).¹⁶ Dado el carácter aleatorio de los componentes del consumo, ε , la probabilidad de que un hogar sea pobre o que no lo sea —teniendo en cuenta sus características observadas— nunca es nula.

16 Este hecho ha sido observado al analizar los errores de las medidas individuales de bienestar derivados de la desigualdad en el seno de los hogares (Haddad y Kanbur, 1990). Véase también Ravallion (1988). En el Instituto Nacional de Estadística e Informática (INEI) del Perú (1996) se ha desarrollado un modelo muy similar al que hemos descrito, pero derivado de estimaciones directas de incidencia y no de las predicciones de las probabilidades de pobreza.

En el cuadro 5 aparecen las estimaciones de incidencia de la pobreza obtenidas a partir de los datos del censo y utilizando nuestros valores de consumo imputado para las ocho regiones geográficas. Comparamos estas tasas con las obtenidas mediante las ECV de los hogares utilizando las cifras de consumo observadas. Conforme a éstas, la incidencia estimada de pobreza en el Ecuador en su conjunto es de 35%. Las tasas de pobreza calculadas sobre la base del consumo imputado mediante los datos del censo se acercan bastante a los obtenidos en la encuesta. En general, las tasas de pobreza derivadas de la encuesta son ligeramente inferiores que las obtenidas del censo —salvo en las zonas rurales de Oriente, que son las mismas para las dos fuentes de datos. Esto se debe, probablemente, a modificaciones en las variables exógenas sobre las que se basan las regresiones de consumo en el período de cuatro años transcurrido entre el censo de 1990 y la encuesta de medición de los niveles de vida de 1994. Por ejemplo, si observamos el cuadro 5, la disminución de la pobreza es más evidente en la región de la Sierra. Como se muestra en el cuadro 1 del apéndice, en esta región la media del número de años de escolaridad del jefe del hogar parece haber aumentado abruptamente entre 1990 y 1994. En el plano regional, los errores estándar de las tasas de pobreza calculadas con datos del censo son sorprendentemente bajas.¹⁷ La clasificación jerárquica de las ocho regiones no es idéntica para las dos fuentes de datos, pero en ambos casos se observa

- 17 Para la incidencia de la pobreza obtenida con datos del censo, el error estándar de nuestro indicador con respecto a la tasa de pobreza real puede calcularse como sigue (para más detalles véase Elbers, Lanjouw y Lanjouw, 1999):

$$P^* = \sum_{i=1}^N \frac{m_i}{M} \left[\frac{\ln z - X_i' \beta}{\hat{\sigma}} \right]$$

$$\text{Var}(P^*) = \left[\frac{\partial P^*}{\partial \beta} \right] \text{Var}(\hat{\beta}) \left[\frac{\partial P^*}{\partial \beta} \right]' + \left[\frac{\partial P^*}{\partial \hat{\sigma}^2} \right]^2 \frac{2\hat{\sigma}^4}{n-k-1} + \sum_{i=1}^N \frac{m_i^2 P_i^* (1-P_i^*)}{M^2}$$

donde n es el tamaño de la muestra del modelo de consumo con k parámetros, estimados utilizando la encuesta del estudio de medición de los niveles de vida. N es el número de hogares que integran la población del censo en la región de interés, m_i es el número de personas que integran el hogar i , y M es el número total de personas que integran la población censada.

$$\frac{\partial P^*}{\partial \hat{\beta}_j} = \sum_{i=1}^N \frac{m_i}{M} \left[\begin{matrix} x_{ij} \\ -\frac{y_j}{\hat{\sigma}} \end{matrix} \right] \phi \left[\frac{\ln z - X_i' \beta}{\hat{\sigma}} \right]$$

donde ij indica el j -ésimo elemento del vector de variables explicativas para el i -ésimo hogar, y

$$\frac{\partial P^*}{\partial \hat{\sigma}^2} = -\frac{1}{2} \sum_{i=1}^N \frac{m_i}{M} \left[\frac{\ln z - X_i' \beta}{\hat{\sigma}^3} \right] \phi \left[\frac{\ln z - X_i' \beta}{\hat{\sigma}} \right]$$

Cuadro 5
TASAS DE POBREZA REGIONALES PARA EL ECUADOR

**Comparación de las tasas obtenidas de la ECV de 1994 con las derivadas
del censo, utilizando los gastos imputados basándose
en un modelo calibrado de la ECV**

Clasificación jerárquica según la ECV	Incidencia de la pobreza (los errores estándar estimados se indican entre paréntesis)		Clasificación jerárquica según el censo
	ECV		
1. Oriente rural	0.67 (0.02)	0.67 (0.004)	(1)
2. Costa rural	0.50 (0.02)	0.52 (0.002)	(3)
3. Sierra rural	0.43 (0.02)	0.53 (0.001)	(2)
4. Guayaquil	0.29 (0.02)	0.35 (0.002)	(4)
5. Quito	0.25 (0.02)	0.33 (0.002)	(5)
6. Costa urbana	0.25 (0.02)	0.29 (0.002)	(6)
7. Oriente urbana	0.20 (0.02)	0.25 (0.009)	(8)
8. Sierra urbana	0.19 (0.02)	0.29 (0.003)	(7)

claramente que las zonas rurales son más pobres que las urbanas y que la zona rural de Oriente es sin duda la más pobre. El Banco Mundial (1996) señaló que, en general, la clasificación jerárquica de las regiones sobre la base de los datos de las encuestas de condiciones de vida no son robustas en el sentido de que a menudo el uso de distintas líneas e índices de pobreza resultan en una reclasificación de las regiones. En tal sentido, la única excepción fue la clasificación jerárquica de las zonas urbanas y rurales que, según pudo comprobarse, era sumamente robusta —se observa una dominancia estocástica de primer orden, según la cual las zonas rurales de Ecuador son consistentemente más pobres que las zonas urbanas. La comparación de las clasificaciones regionales entre los datos de las ECV y del censo es totalmente consistente con estos resultados de dominancia.

Los errores estándar de las tasas de pobreza correspondientes a la encuesta de los niveles de vida incluidas en el cuadro 5 son tales que no podemos rechazar la hipótesis de que las tasas de pobreza en cada sector —urbano y rural, respectivamente— en las distintas regiones son iguales, aunque es posible distinguir estadísticamente el sector urbano del rural. Con los datos del censo, nuestras estimaciones son lo suficientemente precisas para permitir una comparación significativa entre regiones de cada sector.¹⁸

18 Como la comparación de las ocho regiones se basa en distintos modelos de regresión de las encuestas de las condiciones de vida, las estimaciones de los parámetros implícitas en las predicciones del gasto son independientes entre regiones. En este caso, es posible realizar pruebas de significancia estadística de la diferencia de las tasas de pobreza entre la región r y la región u utilizando la fórmula:

$$\text{Var}(P_r^* - P_u^*) = \text{Var}(P_r^*) + \text{Var}(P_u^*).$$

IV. POBREZA EN LAS PROVINCIAS DE ECUADOR. UN EJEMPLO

La metodología esbozada en las secciones anteriores tiene por objeto permitir la construcción de un mapa de pobreza sobre la base de los gastos de consumo en un nivel de desagregación inferior a las ocho grandes regiones para las cuales es válida la ECV. Por ejemplo, en Ecuador hay cerca de 400 cantones, cada uno de los cuales posee cierto grado de autonomía y administración local. A su vez, estos cantones pueden dividirse en un total de más de 1 000 parroquias. Trabajando con los datos del censo se podría calcular fácilmente las tasas esperadas de pobreza a nivel de cantón o de parroquia para determinar en qué lugares se concentra la pobreza. De hecho, como hemos visto en el ejemplo expuesto en la sección II, la información del censo puede utilizarse, en principio, para identificar los hogares pobres y focalizar las transferencias a éstos.

Sin embargo, los errores estándar de las estimaciones de pobreza son una función del grado de desagregación del mapa de pobreza (véase el último término de la primera ecuación de la nota 17. Esto nos advierte sobre los riesgos que entraña tratar de emplear esta metodología para identificar, digamos por caso, los hogares pobres en forma individual.¹⁹ Además, a las objeciones mencionadas es preciso agregar los argumentos bien conocidos que se oponen a éste método de focalización, que se centran en el efecto que podrían tener estas políticas sobre el comportamiento de los beneficiarios potenciales.²⁰

Pese a la necesidad de ser cautos en el uso de la microfocalización, puede ser conveniente elaborar un mapa de pobreza con un nivel de desagregación inferior a las grandes regiones. En última instancia, el nivel óptimo de desagregación dependerá de varios factores. Uno de ellos es su propósito, por ejemplo, ¿el mapa tiene por objeto identificar las zonas administrativas del gobierno de modo que el nivel de desagregación deseado sea algún gobierno local? ¿O, en cambio, tiene por objeto identificar las aldeas y los barrios pobres a fin

9 Supongamos que la probabilidad de pobreza predicha para un hogar determinado fuera de 48%. Para un único hogar, el límite inferior de la estimación del error estándar de la tasa de pobreza de ese hogar sería: $0.49 \approx \sqrt{[0.48(1-0.48)]}$.

10 Van de Walle y Nead (1995) realizan un análisis claro y detallado sobre este tema.

de mejorar la focalización de los proyectos comunitarios (como la infraestructura pública)? También es importante determinar si las estimaciones de los parámetros de un modelo de regresión estimado, digamos por ejemplo, en el plano regional, pueden aplicarse a los desgloses subregionales. En todo este proceso estamos suponiendo implícitamente que en una región determinada el modelo de consumo es el mismo para todo los hogares, independientemente de la provincia, la circunscripción o la comunidad en que estén ubicados.²¹ Esta hipótesis no puede verificarse, aunque es ciertamente menos razonable cuando los niveles de desagregación son muy finos. El nivel deseado de desagregación dependerá, también, de la disponibilidad de otras fuentes de información sobre la pobreza que podrían obtenerse a nivel local. Por último, a determinados niveles de desagregación podrían adquirir mayor importancia y eficacia otros métodos de focalización local, como la autofocalización. De este modo, es probable que la elaboración de un mapa de pobreza sea un proceso secuencial de desagregación gradual que puede continuarse hasta que un nivel de desagregación adicional no aporte nuevos conocimientos. En cada una de las etapas es sumamente importante tener presente cuál es el objetivo del mapa de pobreza.

En el cuadro 6 presentamos los porcentajes de incidencia de la pobreza en Ecuador, desglosadas por provincia, distinguiendo en cada una de ellas las zonas urbanas y rurales. De los datos del cuadro 6 surge claramente que las tasas de pobreza de las distintas provincias varían considerablemente. También podemos ver que, a nivel de provincia, los errores estándar de las tasas de pobreza son bajos. Esto significa que la desagregación a ese nivel no genera una pérdida significativa de precisión en términos estadísticos. En realidad, para que los errores estándar aumenten significativamente en razón del tamaño reducido de las poblaciones será necesario llegar a la elaboración de mapas de pobreza con un alto grado de desagregación espacial. Esto queda

21 Esto depende, en parte, del significado del término ϵ . Es decir, si se lo considera como un efecto fijo para todo los hogares o si el grueso de las variaciones se deben a "shocks" idiosincrásicos sobre los ingresos. Suponemos que la media de ϵ , es cero a ese nivel de estimación. Sin embargo, al trasladarse a los subgrupos, esta hipótesis dejará de ser válida si ϵ , es un efecto fijo para todo los hogares. En este caso, los hogares de un subgrupo determinado pueden tener ingresos relativamente elevados —en función de sus características observables— en comparación con otro subgrupo de características similares. En este caso, es probable que la medida esperada de la pobreza esté sesgada, subestimando el bienestar del primer grupo y sobreestimando el del segundo.

demostrado en el gráfico 1, en el que se representan los errores estándar asociados con los porcentajes de pobreza para las aproximadamente mil parroquias de Ecuador en función de su población. Recién cuando la población de las parroquias es inferior a 500 hogares, el error estándar correspondiente aumentará a niveles tales que no será posible realizar comparaciones confiables.²²

- 22 El error estándar de la diferencia entre las tasas de pobreza de dos parroquias pertenecientes a regiones distintas se calcularía como ya se ha descrito. Sin embargo, como las estimaciones de los parámetros que determinan el gasto imputado son iguales para todas las parroquias de una región determinada, el error estándar de la diferencia de la incidencia de la pobreza entre dos parroquias de una misma región es

$$Var(P_1^* - P_2^*) \approx \left[\frac{\partial(P_1^* - P_2^*)}{\partial \hat{\beta}} \right]' Var(\hat{\beta}) \left[\frac{\partial(P_1^* - P_2^*)}{\partial \hat{\beta}} \right] + \left[\frac{\partial(P_1^* - P_2^*)}{\partial \hat{\sigma}^2} \right]^2$$

$$\frac{2\hat{\sigma}^4}{n-k-1} + \sum_{i=1}^{N_1} \frac{m_i^2 P_{1i}^* (1-P_{1i}^*)}{M_1^2} + \sum_{k=1}^{N_2} \frac{m_k^2 P_{1k}^* (1-P_{1k}^*)}{M_2^2}$$

$$\frac{\partial(P_1^* - P_2^*)}{\partial \hat{\beta}_j} = \sum_{i=1}^{N_1} \frac{m_i}{M_1} \left[-\frac{x_{ij}}{\hat{\sigma}} \right] \phi \left[\frac{\ln z - X'_i \hat{\beta}}{\hat{\sigma}} \right] - \sum_{k=1}^{N_2} \frac{m_k}{M_2} \left[-\frac{x_{kj}}{\hat{\sigma}} \right] \phi \left[\frac{\ln z - X'_k \hat{\beta}}{\hat{\sigma}} \right]$$

donde N , M y m se definen de la misma manera que en la nota 17 para las parroquias 1 y 2, cuyos subíndices son, respectivamente, i y h , y donde j es el j -ésimo elemento del vector dado.

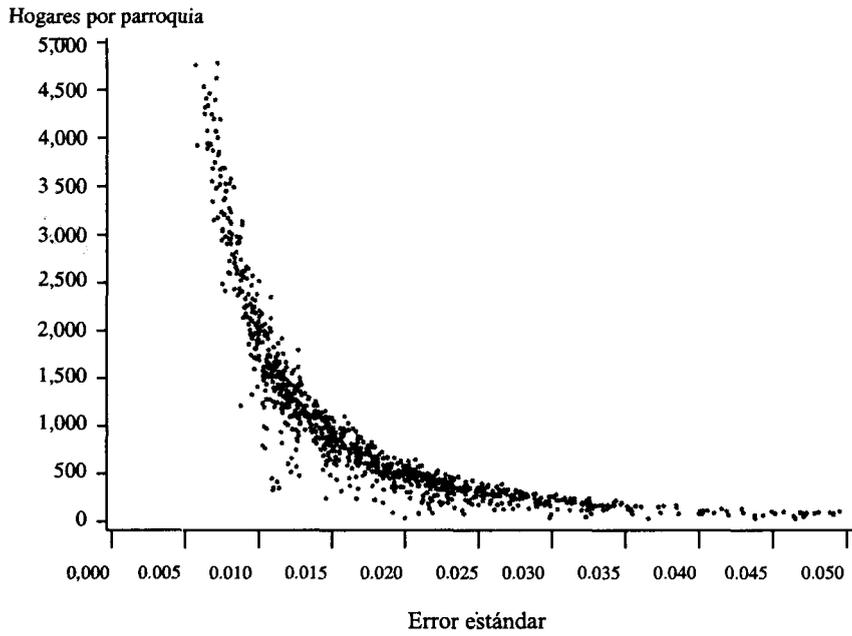
$$\frac{\partial(P_1^* - P_2^*)}{\partial \hat{\sigma}^2} = -\frac{1}{2} \sum_{i=1}^{N_1} \frac{m_i}{M_1} \left[\frac{\ln z - X'_i \hat{\beta}}{\hat{\sigma}^3} \right] \phi \left[\frac{\ln z - X'_i \hat{\beta}}{\hat{\sigma}} \right] - \frac{1}{2} \sum_{k=1}^{N_2} \frac{m_k}{M_2} \left[\frac{\ln z - X'_k \hat{\beta}}{\hat{\sigma}^3} \right] \phi \left[\frac{\ln z - X'_k \hat{\beta}}{\hat{\sigma}} \right]$$

Cuadro 6

MAPA DE POBREZA DE ECUADOR: PROVINCIAS URBANAS Y RURALES

Región	Sector	Provincia	Tasa de po- breza esperada	Error estándar
Oriente	Rural		0.67	0.004
		Pastaza	0.65	0.005
		Sucumbios	0.65	0.005
		Morona Santiago	0.66	0.005
		Zamora Chinchipe	0.67	0.005
		Napo	0.69	0.004
Sierra	Rural		0.53	0.001
		Tungurahua	0.45	0.002
		Pichincha	0.46	0.002
		Azuay	0.50	0.002
		Canar	0.52	0.003
		Bolívar	0.55	0.003
		Imbabura	0.56	0.003
		Loja	0.57	0.003
		Carchi	0.58	0.004
		Chimborazo	0.59	0.003
		Cotopaxí	0.63	0.003
Costa	Rural		0.52	0.002
		El Oro	0.45	0.003
		Guayas	0.48	0.002
		Los Ríos	0.55	0.002
		Manabi	0.56	0.002
		Esmeraldas	0.59	0.003
		Galápagos	0.14	0.008
Costa	Urbana	Guayaquil	0.35	0.002
Sierra	Urbana	Quito	0.33	0.002
Costa	Otras zonas urbanas		0.29	0.002
		El Oro	0.24	0.003
		Esmeraldas	0.27	0.004
		Manabi	0.29	0.003
		Guayas	0.30	0.003
		Los Ríos	0.32	0.003
Sierra	Otras zonas urbanas		0.29	0.003
		Azuay	0.23	0.003
		Tungurahua	0.25	0.004
		Chimborazo	0.25	0.004
		Cotopaxí	0.28	0.004
		Loja	0.31	0.004
		Canar	0.31	0.006
		Imbabura	0.33	0.004
		Carchi	0.33	0.005
		Pichincha	0.33	0.003
Oriente	Urbana		0.25	0.009
		Pastaza	0.24	0.011
		Zamora Chinchipe	0.24	0.013
		Morona Santiago	0.28	0.013

Gráfico 1
**ERRORES DEL PORCENTAJE DE POBREZA
SEGÚN NIVEL DE DESAGREGACIÓN**
Estimación a nivel de parroquias



V. OBSERVACIONES FINALES

En muchos países en desarrollo los mapas de pobreza desempeñan un papel importante para orientar la asignación del gasto público dirigido a reducir la pobreza. En lo esencial, estos mapas son un perfil geográfico de la pobreza, que indican en qué zonas del país se concentra este fenómeno y, por ende, en cuáles es dable esperar que las políticas pertinentes tendrán un impacto más importante. Los mapas de la pobreza son tanto más útiles cuanto más refinado sea el nivel de desagregación.

Para lograr esos niveles de desagregación, es fundamental poder trabajar con conjuntos de datos de amplia cobertura. Sin embargo, es infrecuente encontrar datos de encuestas que tengan esa característica y que al mismo tiempo incluyan información detallada sobre el bienestar de los hogares. En general, se adopta una solución de compromiso entre el tamaño de la muestra y la calidad porque lograr ambos objetivos resulta muy caro, tanto en términos financieros como administrativos.

En el presente estudio hemos analizado las posibilidades de combinar lo mejor de ambas fuentes de datos a fin de construir un mapa de pobreza desagregado que tenga en cuenta al mismo tiempo medidas de bienestar en función de los ingresos y del consumo. Primeramente, demostramos que la elaboración de un mapa de pobreza sobre la base de datos censales pero utilizando un esquema de ponderación arbitrario puede no ser adecuado para focalizar los hogares pobres en términos del nivel de consumo. Es posible que los programas de transferencia para reducir la pobreza elaborados utilizando mapas de ese tipo sólo lleguen a un subconjunto de los beneficiarios previstos y que una proporción significativa de los fondos se desvíen a personas que no son pobres.

Nosotros proponemos un enfoque alternativo: utilizando datos sobre los hogares del Ecuador, obtenidos mediante una encuesta de condiciones de vida de pequeño tamaño muestral pero de alta calidad (ECV, 1994), modelamos directamente el consumo como función de variables explicativas incluidas también en el censo. Como resulta posible explicar buena parte de las variaciones del consumo de los hogares observadas en la ECV mediante unas pocas variables explicativas incluidas tanto en el censo como en la ECV, la incidencia de la pobreza calculada a partir de los datos del censo —sobre la base de esta cifra de consumo imputada— es muy cercano al obtenido mediante la ECV. También demostramos que en el Ecuador las tasas de pobreza calculadas mediante los datos del censo generalmente son muy

precisas en términos estadísticos. Esta precisión disminuye inevitablemente a medida que se aumenta el grado de desagregación espacial. Aunque podría resultar muy tentador usar la metodología desarrollada en este estudio para identificar en forma individual a los hogares pobres, hemos demostrado que esta aplicación sería muy inadecuada. Sostenemos que el enfoque desarrollado en el presente estudio puede utilizarse hasta un grado relativamente alto de desagregación, pero debería complementarse con otras fuentes de información e investigaciones adicionales.

Es probable que una de las aplicaciones prácticas más útiles de esta metodología sea la comparación de patrones regionales de otros indicadores de bienestar, oportunidades y acceso. Por ejemplo, se podría superponer un mapa de patrones de acceso a los centros de atención primaria de salud y un mapa elaborado con nuestra metodología, en el que se indique las zonas en que se concentra la pobreza. Este tipo de metodología podría resultar sumamente útil para los encargados de formular las políticas, por varios motivos. Podría ayudar a éstos a determinar en qué zonas es preciso concentrar los esfuerzos dirigidos a ampliar el acceso a los centros de atención primaria de salud. También podría contribuir a analizar de qué manera se puede ampliar el acceso a la atención primaria de salud —quizás se desee subsidiar el acceso en las zonas pobres y poner a prueba métodos de recuperación de costos en las zonas menos pobres. Asimismo, una correlación estrecha entre, digamos por caso, los patrones regionales de pobreza rural y el acceso a las carreteras, también puede dar algunas pistas sobre las *causas* posibles de la pobreza. Este tipo de metodología podría utilizarse para una amplia gama de indicadores: niveles de salud y de educación; pertenencia a distintas etnias o grupos indígenas; acceso a la infraestructura y otros servicios públicos; calidad de la tierra y ecología; medio ambiente, etc.

Finalmente, como se señala en la introducción, la capacidad de elaborar mapas de pobreza con un alto grado de desagregación también puede contribuir a resolver problemas de investigación más amplios. Una posibilidad es analizar la variación espacial de la relación entre la desigualdad y el desempeño económico en un país determinado, de manera análoga a los análisis entre países, tema que está despertando mucho interés. Mediante este enfoque es muy probable que se pueda evitar algunos de los problemas metodológicos que plantea el análisis entre países. También permitiría abordar otros problemas de investigación. Por ejemplo, la idea de que las propias comunidades son

las más indicadas para identificar el tipo de intervenciones más beneficiosas para los pobres que las integran esta implícita en algunos de los argumentos actuales en favor de la descentralización de las actividades de reducción de la pobreza. En alguna medida, esta posición se apoya en el argumento de que a nivel comunitario es menos probable que un subconjunto de hogares que no son pobres se apropie de los recursos públicos. Probablemente, ello está vinculado a los niveles de desigualdad existentes en las comunidades, un aspecto que nunca ha resultado fácil investigar. Con la metodología presentada en este trabajo, se podría analizar el consumo de los hogares deducido de los censos a fin de evaluar el grado de desigualdad en las zonas geográficas más pequeñas.

BIBLIOGRAFÍA

- Alesina, Alberto y Dani Rodrik (1994), "Distributive policies and economic growth", *The Quarterly Journal of Economics*, N° 109.
- Angrist, J.D. y A.B. Krueger (1992), "The effect of age of school entry on educational attainment: an application of instrumental variables with moments from two samples", *Journal of the American Statistical Association*, N° 87.
- Arellano, M. y C. Meghir (1992), "Female labour supply and on the job search: an empirical model estimated using complementary data sets", *Review of Economic Studies*, N° 59.
- Atkinson, A. (1989), *Poverty and Social Security*, Hemel Hempstead, Harvester Wheatsheaf.
- Banco Mundial (1996), "Ecuador Poverty Report", World Bank Country Study, Washington, D.C.
- Bramley, G. y G. Smart (1996), "Modelling local income distributions in Britain", *Regional Studies*, N° 30.
- Bruno, Michael, Martin Ravallion y Lyn Squire (1998), "Equity and growth in developing countries: old and new perspectives on the policy issues", *Income Distribution and High Quality Growth*, Vito Tanzi y Ke-Young Chu (comps.), Cambridge, Massachusetts, The MIT Press.
- Chaudhuri, Shubham y Martin Ravallion (1994), "How well do static welfare indicators identify the chronically poor?", *Journal of Public Economics*, vol. 53, N° 3.

- Deninger, Klaus y Lyn Squire (1996), "A new data set measuring income inequality", *The World Bank Economic Review*, vol. 10, N° 3, Washington, D.C., Banco Mundial.
- Elbers, C., Jean Olson Lanjouw y Peter Lanjouw (1999), "Welfare in Villages and Towns: Micro-measurement of Poverty and Inequality", Amsterdam, Universidad Libre de Amsterdam, inédito.
- Farrell, P., B. MacGibbon y T.J. Tomberlin (1997), "Empirical bayes estimation using logistic regression models and summary statistics", *Journal of Business and Economic Statistics*, N° 15.
- Fields, B. (1989), "Changes in poverty and inequality in developing countries", *World Bank Research Observer*, N° 4.
- FONCODES (Fondo Nacional de Compensación y Desarrollo Social) (1995), *Mapa de pobreza en Perú*, Lima.
- Gobierno de El Salvador (1995), "Priorización de municipios a partir de datos censales", San Salvador, Dirección General de Política Económica y Social, Ministerio de Coordinación del Desarrollo Económico y Social.
- Haddad, L. y R. Kanbur (1990), "How serious is the neglect of intra-household inequality?", *The Economic Journal*, N° 100.
- Hentschel, Jesko y Peter Lanjouw (1996), "Constructing an Indicator of Consumption for the Analysis of Poverty: Principles and Illustrations with Reference to Ecuador", LSMS Working Paper, N° 124, Washington, D.C., Departamento de Investigaciones sobre Políticas de Desarrollo, Banco Mundial.
- ____ (1995), "Perfil de la pobreza en Ecuador", *Cuestiones económicas*, Quito, Banco Central del Ecuador.
- Imbens, G. y J. Hellerstein (1998), "Imposing Moment Restrictions from Auxiliary Data by Weighting", NBER Technical Working Paper, N° 202.
- INEI (Instituto Nacional de Estadística e Informática) (1996), *Metodología para determinar el ingreso y la proporción de hogares pobres*, Lima.
- Isaki, C.T. (1990), "Small-area estimation of economic statistics", *Journal of Business and Economic Statistics*, N° 8.
- Lanjouw, Jean y Peter Lanjouw (1997), "Poverty Comparisons with Noncompatible Data: Theory and Illustrations", Policy Research Working Paper, N° 1709, Washington, D.C., Departamento de Investigaciones sobre Políticas de Desarrollo, Banco Mundial.

- Lusardi, A. (1996), "Permanent income, current income and consumption: evidence from two panel data sets", *Journal of Business and Economic Statistics*, vol. 14, N° 1.
- Paulin, G.D. y D.L. Ferraro (1994), "Imputing income in the consumer expenditure survey", *Monthly Labor Review*, December.
- Purcell, N.J. y L. Kish (1980), "Postcensal estimates for local areas (or domains)", *International Statistical Review*, N° 48.
- Ravallion, Martin (1994), *Poverty Comparisons*, Chur, Harwood Academic Publishers.
- _____(1988), "Expected poverty under risk-induced welfare variability", *The Economic Journal*, N° 98.
- Sen, A.K. (1984), *Resources, Values and Developmentk*, Oxford, Blackwell.
- van de Walle, D. y K. Nead (comps.) (1995), *Public Spending and the Poor: Theory and Evidence*, Baltimore, Johns Hopkins University Press.

APÉNDICE

Cuadro 1

ESTADÍSTICAS DESCRIPTIVAS COMPARADAS

Algunas variables obtenidas en la encuesta de medición de niveles de vida de 1994 y las del censo de 1990

		Sierra rural		Sierra urbana	
		EMNV*	Censo	EMNV*	Censo
Años de escolaridad del jefe del hogar	media (d.s.)	4.48 (3.36)	4.33 (4.18)	8.75 (5.15)	8.11 (5.19)
El jefe del hogar es hombre	media	0.84	0.78	0.81	0.77
Personas por dormitorio	media (d.s.)	3.04 (1.78)	3.28 (2.05)	2.42 (1.49)	2.59 (1.66)
Conexión a la red pública de suministro de agua potable	media	0.31	0.52	0.94	0.89
Recolección de residuos por camiones	media	0.25	0.19	0.80	0.81
Retrete con descarga de agua	media	0.37	0.24	0.69	0.68
Conexión telefónica	media	0.04	0.07	0.31	0.27
		Quito		Costa rural	
		EMNV*	Censo	EMNV*	Censo
Años de escolaridad del jefe del hogar	media (d.s.)	10.67 (5.25)	9.52 (5.23)	3.63 (3.16)	4.12 (4.01)
El jefe del hogar es hombre	media	0.82	0.79	0.96	0.87
Personas por dormitorio	media (d.s.)	2.21 (1.32)	2.45 (1.52)	3.74 (1.98)	3.73 (2.29)
Conexión a la red pública de suministro de agua potable	media	0.90	0.83	0.08	0.21
Recolección de residuos por camiones	media	0.89	0.88	0.05	0.12
Retrete con descarga de agua	media	0.79	0.68	0.27	0.33
Conexión telefónica	media	0.43	0.36	0.00	0.03

Cuadro 1 Apéndice (conclusión)

	EMNV*	Costa urbana		Guayaquil	
		Censo	EMNV*	Censo	
Años de escolaridad del jefe del hogar	media (d.s)	6.64 (4.42)	6.91 (4.98)	8.88 (4.89)	8.65 (4.96)
El jefe del hogar es hombre	media	0.83	0.80	0.77	0.78
Personas por dormitorio	media (d.s)	3.16 (2.08)	3.12 (1.96)	3.01 (1.87)	2.99 (1.92)
Conexión a la red pública de suministro de agua potable	media	0.55	0.71	0.72	0.62
Recolección de residuos por camiones	media	0.57	0.56	0.75	0.54
Retrete con descarga de agua	media	0.66	0.71	0.76	0.75
Conexión telefónica	media	0.12	0.13	0.25	0.23
		Oriente rural		Oriente urbano	
		EMNV*	Censo	EMNV*	Censo
Años de escolaridad del jefe del hogar	media (d.s)	5.82 (3.94)	5.16 (4.08)	8.27 (4.50)	8.29 (4.7)
El jefe del hogar es hombre	media	0.89	0.85	0.83	0.78
Personas por dormitorio	media (d.s)	3.54 (1.89)	3.49 (2.26)	2.54 (1.50)	2.64 (1.63)
Conexión a la red pública de suministro de agua potable	media	0.16	0.29	0.92	0.87
Recolección de residuos por camiones	media	0.10	0.20	0.81	0.83
Retrete con descarga de agua	media	0.23	0.18	0.66	0.60
Conexión telefónica	media	0.01	0.03	0.24	0.14

* EMNV = estudio de medición de los niveles de vida;
d.s. = desviación estándar.

INDICADORES ALTERNATIVOS PARA FOCALIZAR EL GASTO SOCIAL EN PARAGUAY*

Marcos Robles

Consultor del Programa para el Mejoramiento
de las Encuestas y la Medición de las Condiciones
de Vida en América Latina y el Caribe
(MECOVI-Paraguay)

RESUMEN

El propósito de la presente investigación es estimar el porcentaje de población que no logra adquirir una canasta básica de consumo por distritos y departamentos. Se utilizó información combinada de la Encuesta Integrada de Hogares (EIH) de 1997/1998 y el último Censo Nacional de Población y Viviendas (CNPV), así como técnicas econométricas e instrumentos analíticos recientemente desarrollados para este propósito.

Las estimaciones tienen, a nivel de distritos, la representatividad de la información censal (de la que carece una encuesta), y parte de la riqueza de datos sobre los niveles de vida de la población que captan las encuestas (lo que no hace un censo). De este modo, la información provista por los mapas de pobreza basados en los ingresos y el consumo de la población pueden servir como herramientas para diseñar políticas eficientes que busquen mejorar la capacidad adquisitiva de los hogares pobres de Paraguay.

* El autor agradece de manera especial la colaboración del Departamento de Encuestas y el Departamento de Cómputo de la Dirección General de Estadística, Encuestas y Censos, y a Jesko Hentschel, del Banco Mundial, por su asesoramiento "en línea" desde el inicio de la investigación.

ALTERNATIVE INDICATORS FOR TARGETING SOCIAL SPENDING IN PARAGUAY

ABSTRACT

The aim of this research project is to estimate the percentages of the population, by district and department, that cannot afford a specified basket of staple consumer goods. The study draws on information from the 1997/1998 Integrated Household Survey and the most recent National Population and Housing Census, together with econometric techniques and analytical tools that have recently been developed for this purpose.

The estimates prepared at the district level are as representative as census information is (i.e., more representative than survey data) and are based on the wide range of data regarding the population's living standards that are gathered in surveys (something which censuses cannot do). Hence, the information provided by poverty maps based on income and consumption patterns can be used as a tool to design efficient policies for increasing the purchasing power of poor households in Paraguay.

I. INTRODUCCIÓN

En los últimos años, para diagnosticar la situación de la pobreza en Paraguay e identificar a los grupos más deprimidos se han utilizado diversas fuentes de datos y métodos de medición. Por una parte, con las encuestas de hogares llevadas a cabo por la Dirección General de Estadística, Encuestas y Censos (DGEEC), y mediante el método de medición conocido como de línea de pobreza, se ha determinado el nivel y también el perfil sociodemográfico de la población en situación de pobreza. De acuerdo con este método, son pobres quienes viven en hogares cuyo ingreso o consumo per cápita es inferior al costo de una canasta básica de consumo, por lo que se recomienda aplicar políticas de tipo salarial, de empleo y de generación de ingresos.

Por otra parte, con el método de medición denominado de necesidades básicas insatisfechas (NBI), y sobre la base de los datos de los censos nacionales de población y viviendas, se han construido mapas de pobreza. Éstos han permitido jerarquizar las diferentes divisiones geográficas del país con arreglo a la proporción de población que no alcanza un nivel mínimo de satisfacción en algunos aspectos básicos relacionados con las condiciones de vivienda, educación y salud, entre otros, lo que hace recomendable ejecutar acciones específicas, como establecer sistemas de crédito para vivienda, proveer servicios de agua y eliminación de excretas, realizar programas de educación y otras similares.

Esto significa que, dependiendo de la información obtenida con cada método de medición, las políticas gubernamentales para reducir la pobreza se han definido con aproximaciones y alcances diferentes. Desde el punto de vista de la focalización del gasto social, sin embargo, sólo se ha contado con datos recopilados con un método de medición, lo cual ha impedido contar con una visión más amplia del carácter multidimensional de la pobreza y, en consecuencia, poder definir políticas sociales más eficientes.

El propósito de la presente investigación es llenar este vacío mediante la estimación distrital y departamental del ingreso y consumo familiares, lo que permitiría determinar la proporción de hogares que no logra adquirir una canasta básica de consumo. Para tal efecto se ha utilizado información procedente tanto de la Encuesta Integrada de Hogares (EIH) de 1997/1998 como del último Censo Nacional de Población y Viviendas (CNPV), a la que se le aplicaron técnicas econométricas e instrumentos analíticos recientemente desarrollados para este propósito (Hentschel y otros, 1999).

De este modo se ha podido llegar a estimaciones que, a nivel de distritos, tienen la representatividad de la información censal (lo que no sucede en el caso de la EIH) y parte de la riqueza de datos sobre los niveles de vida de la población que proveen las encuestas de hogares (de la que carece el CNPV). Como se sabe, los resultados del Censo no permiten estimar directamente la proporción de hogares por debajo de la línea de pobreza, un indicador sintético bastante aceptado y utilizado, debido a que no se indagó sobre los ingresos o gastos familiares. Las encuestas de hogares, por su parte, si bien contienen información bastante detallada respecto de los ingresos y gastos familiares, así como de numerosos otros indicadores de bienestar, su representatividad geográfica no es suficiente como para satisfacer la demanda de datos con fines de focalización.

En este informe también se presentan indicadores alternativos de focalización geográfica, basados en las estimaciones efectuadas y en la información que publica la DGEEC, que luego se evalúan con respecto a conceptos conocidos, como errores de exclusión, inclusión, filtración y subcobertura de los programas sociales, y también a los efectos que ejercen sobre los niveles de pobreza las transferencias simuladas a los hogares.

Cabe señalar finalmente que, en el contexto de la presente investigación, las estimaciones elaboradas con datos de la EIH se basaron en información válida sobre 20 664 personas entrevistadas, correspondientes a 4 353 hogares; cuando se utilizaron datos del CNPV, la información que se procesó comprende a 4 111 991 personas censadas, correspondientes a 863 990 hogares. La DGEEC ejecutó la EIH entre agosto de 1997 y julio de 1998, y el CNPV el 26 de agosto de 1992.

1. ¿Por qué estimar tasas de pobreza desagregadas sobre la base del ingreso y el consumo familiares?

La respuesta es que existen varias razones, pero la más importante es la posibilidad de observar la heterogeneidad geográfica de la pobreza con un enfoque que no sea el de las NBI; es decir, abordar con un criterio adicional las múltiples dimensiones de la pobreza y, en consecuencia, definir políticas más eficientes para combatirla.

Con el enfoque de las NBI sólo se ha podido describir de manera desagregada algunas facetas del bienestar de la población, a través de indicadores construidos con información procedente de los censos de población y vivienda. Así, sobre la base del censo de 1982 se elaboró (en 1991) un mapa de pobreza desagregado al nivel de áreas urbanas y rurales de cada departamento, en el que se consideraban tres indicadores de NBI: calidad de la vivienda, hacinamiento e infraestructura sanitaria. Con el censo de 1992 se elaboró (en 1995) un mapa a nivel de áreas urbanas y rurales de cada distrito y cuatro indicadores de NBI: calidad de la vivienda, infraestructura sanitaria, acceso a la educación y capacidad de subsistencia (STP/DGEEC, 1991; STP/DGEEC/FNUAP/PNUD, 1995). Estas mediciones se han visto limitadas por la cantidad de información que es posible captar de estas fuentes, ya que los censos se ejecutan en un solo día y con cuestionarios muy concisos, dado el alto costo que supone entrevistar a toda la población.

Mediante la estimación de los niveles de pobreza a nivel de distritos y departamentos, en la que se considere la capacidad de los hogares (medida a través de sus gastos de consumo e ingresos) para acceder a una canasta básica de consumo, podrá conocerse en forma desagregada la dimensión económica (o monetaria) del bienestar de la población, medida que en la bibliografía sobre el tema es considerada como una de las más confiables para estimar la pobreza.¹ Se tendrá entonces no sólo un criterio adicional para analizar la pobreza, sino también información desagregada exenta de los siguientes problemas que muestra el método de medición de las NBI:

- i) El requerimiento implícito de hacer que todas las necesidades básicas sean iguales, debido a que para calcular el porcentaje de población con al menos una NBI se toma en cuenta si tiene una

1 Véase, entre otros, Glewwe y van der Gaag (1988).

NBI o no, pero sin establecer previamente una categorización o tipificación de estas necesidades.

- ii) El énfasis en las características de la vivienda (dos de los cuatro indicadores se refieren a ellas), en tanto que necesidades quizás más relevantes, como de nutrición y de salud, reciben poca atención o se miden imperfectamente.²
- iii) La omisión de medidas que muestren la profundidad de la pobreza, ya que se incluye en la misma categoría a individuos que tienen una o cuatro NBI; esto sucede incluso respecto de un mismo indicador, ya que se considera igualmente pobres a individuos que viven en hogares con tres, cuatro o cinco miembros por dormitorio.

En consecuencia, la construcción de un mapa de pobreza basado en los ingresos y el consumo de la población puede constituirse en una herramienta importante para formular políticas eficientes que busquen una mejoría de la capacidad adquisitiva de los hogares pobres de Paraguay, o para diseñar esquemas de intervención y localización de transferencias o subsidios. Se sabe que los recursos se usan mejor si los grupos necesitados pueden ser reconocidos con mayor precisión. Es decir, si el gasto social se orientara en función de un mapa de pobreza detallado sería posible reducir los errores de exclusión (pobres no beneficiados con respecto a la población total) e inclusión (no pobres beneficiados con respecto a la población total) de los programas sociales, como se verá más adelante.

Los mapas de pobreza son herramientas que facilitan y hacen más comprensible la presentación de información referida a las tendencias espaciales, agrupamientos y concentraciones de la pobreza. Por consiguiente, su utilidad rebasa el ámbito del gobierno y de las organizaciones no gubernamentales, para alcanzar, en especial, el de las comunidades locales y regionales.

Otra razón, también importante, para construir mapas con criterios monetarios, es la posibilidad de combinarlos con otras fuentes de información a fin de capturar aspectos adicionales de la pobreza y la

2 En el marco de este método se supone que el indicador de infraestructura sanitaria aproxima al indicador de salud; sin embargo, según la EIH de 1997/1998, entre el indicador de NBI referidas a infraestructura sanitaria y la proporción de enfermos en el hogar existe sólo una correlación positiva de 6% y entre dicho indicador de NBI y la proporción de enfermos que acude al servicio médico, una correlación negativa de sólo 3%. Del mismo modo, se supone que el indicador de capacidad de subsistencia aproxima al indicador de capacidad de ingreso y consumo de la población, no obstante lo cual entre el indicador de NBI referidas a capacidad de subsistencia y el consumo per cápita existe una correlación negativa de 13% y entre ese mismo indicador y el ingreso per cápita, una correlación de 15%.

desigualdad que no son fáciles de medir a través de una encuesta o censo (Deichmann, 1999). Por ejemplo, la información sobre pobreza medida por ingresos combinada con datos de predicción climática sería fundamental para determinar el grado de vulnerabilidad de la producción agropecuaria y algunos aspectos relacionados con la seguridad alimentaria del país, dada la alta dependencia de dicha producción de los periodos de lluvia, sequía, tormentas y otros.³

Otra información que puede combinarse con los mapas de pobreza es la referida al acceso de la población a mercados y servicios. Por ejemplo, podría construirse un indicador que midiera la proporción de población que vive cerca o lejos de los centros educativos o puestos de salud (indicador de equidad en el acceso a los servicios básicos) y tener, junto con la información sobre pobreza, datos que indicaran al gobierno no solo hacia dónde dirigir los esfuerzos y recursos para construir infraestructura básica que beneficie preferentemente a la población vulnerable,⁴ sino también cómo implementarlos. Por ejemplo, si se trata de un proyecto de salud, con la información construida se sabrá dónde es necesario asignar subsidios y dónde implementar esquemas de recuperación de costos (Hentschel y otros, 1999).

2. ¿Cómo estimar las tasas de pobreza distrital y departamental?

El procedimiento utilizado en la presente investigación fue el siguiente: con la información de la EIH se estimaron modelos econométricos de ingreso y consumo; las variables explicatorias elegidas fueron aquellas que también se encuentran en el CNPV. Luego se aplicaron los parámetros estimados de estos modelos a la información del CNPV para predecir la probabilidad de que los hogares (censados) fueran pobres.

Fue necesario realizar las siguientes actividades:

- i) Revisión de los cuestionarios de ambas fuentes para seleccionar las preguntas con formato similar y las opciones de respuesta compatibles, que permitieran construir las variables predictoras del ingreso familiar. En total se encontraron 28 preguntas comunes

3 Instituciones como la Dirección de Meteorología e Hidrografía de la Dirección Nacional de Aeronáutica Civil y el Programa de Agrometeorología del Ministerio de Agricultura y Ganadería podrían facilitar esta información.

4 En Robles (1999b) se ha mostrado econométricamente la importancia de estos servicios para los niveles de bienestar de la población.

respecto de los siguientes temas: **Vivienda:** tipo y condición de ocupación de la vivienda; materiales en paredes exteriores, piso y techo; abastecimiento de agua, alumbrado, y eliminación de basura. **Hogar:** piezas para dormir; pieza y combustible para cocinar; pieza para bañarse; servicio sanitario; equipamiento. **Individuo:** relación de parentesco, sexo, edad, lugar de nacimiento, educación, estado civil, condición de actividad, ocupación principal, categoría de ocupación, rama de actividad. **Ubicación geográfica:** departamento, distrito y área de residencia.

- ii) Recodificación y transformación de las opciones de respuesta a las preguntas seleccionadas, para lo cual se tomó en consideración la posibilidad de que pudieran expresarse en forma cuantitativa, categórica o ambas, que sus efectos sobre el ingreso pudieran ser aditivos o multiplicativos y que su relación con el ingreso pudiera tomar la forma lineal, logarítmica, inversa o cuadrática. De este modo, antes de estimar los modelos, se dispuso de un total de 114 variables comunes a ambas fuentes.
- iii) Regresión del logaritmo natural del ingreso familiar por persona con las variables descritas, mediante el método de mínimos cuadrados ponderados (siendo el ponderador el factor de expansión a nivel de población). En el proceso de estimación se descartaron las variables redundantes y poco significativas, para optar finalmente por las relaciones más simples (aditiva y lineal), si las contribuciones de las variables predictivas fueran parecidas.⁵ Este proceso se ejecutó en forma paralela para las siete áreas geográficas del país representativas a nivel de la EIH: Asunción, departamentos de San Pedro, Caaguazú, Itapúa, Alto Paraná, Central y resto.⁶
- iv) Aplicación de los parámetros de las regresiones estimadas (coeficientes y error estándar), junto con el valor de las líneas de pobreza correspondientes (Robles, 1999a), a los datos censales para obtener los ingresos, consumo y la probabilidad de ser pobre de cada uno de los hogares censados y, por ende, de los diferentes distritos y departamentos del país. En este marco, la proporción de la población pobre en cada espacio geográfico es igual al promedio ponderado (por el tamaño del hogar) de las probabilidades individuales.

5 Para una descripción de las variables que finalmente se utilizaron para estimar los modelos de ingreso y consumo, véase el anexo A de Robles (1999b).

6 Además, en cada una de las áreas geográficas, se excluyeron los casos con comportamiento atípico con respecto a la distribución del ingreso y consumo familiar per cápita.

II. MARCO CONCEPTUAL

El procedimiento descrito puede formalizarse así: el indicador de pobreza para un hogar i (P_i), dado el valor de la línea de pobreza (z), se construye con los siguientes criterios:

Un hogar es pobre si $Lnipc_i < Lnz \Rightarrow P_i = 1$

Un hogar es no pobre si $Lnipc_i \geq Lnz \Rightarrow P_i = 0$

donde ipc_i es el ingreso (o el consumo) per cápita del hogar i y el prefijo Ln indica logaritmo natural. Con ellos, la incidencia de la pobreza (P) para un distrito o departamento podría estimarse de acuerdo con la expresión:

$$P = \frac{1}{\sum_{i=1}^n T_i} \sum_{i=1}^n T_i * P_i$$

donde n es el número de hogares en el distrito o departamento y T_i es el tamaño de la familia.

Al contar con información sobre las características observables del hogar i (X_i), la pobreza esperada (P_i) puede definirse como:⁷

$$E[P_i/X_i, \beta, \sigma] = \Phi[(Lnz - X_i' \beta) / \sigma]$$

donde Φ es una función de distribución normal estándar acumulativa (es decir, la incidencia de la pobreza igual a la probabilidad de que sea menor a $[Lnz - X_i' \beta] / \sigma$).

Para un departamento o distrito, la pobreza esperada, teniendo en cuenta la información de las X_i , será:

$$E[P/X_i, \beta, \sigma] = \frac{1}{\sum_{i=1}^n T_i} \sum_{i=1}^n T_i * \Phi[(Lnz - X_i' \beta) / \sigma]$$

7 El resto de esta sección se basa en Hentschel y otros (2000).

Con los datos de la EIH, los parámetros β y σ pueden estimarse con arreglo al modelo:

$$Lnipc_i = X_i'\beta + \varepsilon_i$$

donde ε_i es un término de perturbación aleatoria que se distribuye normalmente con media cero y varianza σ^2 , y las X_i son variables comunes a la EIH y el CNPV.

Estimados β y σ , y teniendo en consideración el valor de las líneas de pobreza correspondientes, la pobreza esperada del hogar i (del CNPV) y la de un distrito o departamento, podrían definirse, respectivamente, con las siguientes expresiones:

$$E[P_i/X_i, \hat{\beta}, \hat{\sigma}] = \Phi[(Lnz - X_i'\hat{\beta})/\hat{\sigma}]$$

$$E[P/X_i, \hat{\beta}, \hat{\sigma}] = \frac{1}{\sum_{i=1}^n T_i} \sum_{i=1}^n T_i * E[P_i/X_i, \hat{\beta}, \hat{\sigma}]$$

Esto significa que la condición de pobreza de un hogar es estimada sobre la base de la probabilidad de ser pobre, y la de un distrito o departamento, sobre la base de un promedio ponderado (por el tamaño del hogar) de dichas probabilidades (y no como resultado de un conteo de quienes tienen ingresos por debajo de la línea de pobreza).⁸

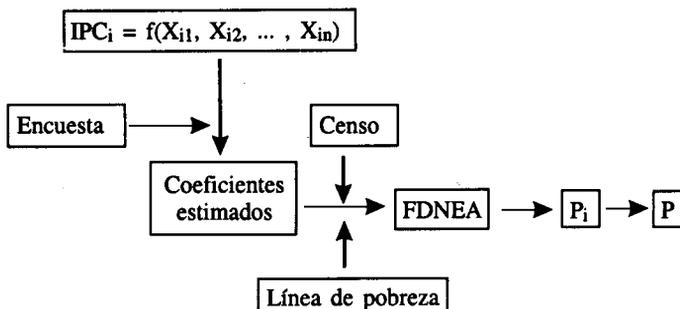
En el diagrama 1 se presenta un resumen de los pasos seguidos dentro de este marco analítico.

8 En forma adicional, con vistas a estudios posteriores sobre la distribución del ingreso y el consumo, así como sobre otras medidas de pobreza a nivel de distritos, y como los modelos no explican perfectamente dichos ingreso y consumo (es decir, R^2 nunca llega a ser igual a 1), éstos se estimaron según la siguiente expresión:

$$i\hat{p}c_i = \exp(Ln\hat{i}pc_i + 0,5*\hat{\sigma}^2)$$

(Véase Demombynes, 1999).

Diagrama 1
CONSTRUCCIÓN DE LAS TASAS DE POBREZA A NIVEL DE DISTRITOS



- IPC_i = ingreso per cápita del hogar i
 X_{ij} = característica socioeconómica j del hogar i (variables existentes en la encuesta y censo)
 P_i = probabilidad de que el hogar i sea pobre dado sus X_i
 P = incidencia de la pobreza de un departamento o distrito (promedio de P_i , ponderado por el tamaño de la población)
 FDNEA = función de distribución normal estandarizada acumulativa

1. Resultados: jerarquización geográfica de la pobreza

La aplicación del procedimiento antes descrito, con la información de la EIH y el CNPV, generó los siguientes resultados:

- i) El poder predictivo de los 14 modelos estimados (aproximados con el valor del coeficiente de determinación) varían entre 47% (San Pedro) y 76% (Alto Paraná) en el caso del ingreso, y entre 63% (San Pedro) y 77% (Alto Paraná) en el del consumo.⁹
- ii) La incidencia de la pobreza calculada con los datos del CNPV es bastante cercana a la obtenida con los de la EIH (véase el cuadro 1).¹⁰

⁹ En el anexo B de Robles (1999b), puede verse el detalle de las estimaciones realizadas sobre los parámetros estimados, los errores estándar y las pruebas de significancia correspondientes a cada uno de los modelos.

¹⁰ Aunque, en general, son un poco más altas debido a los cambios experimentados por las variables independientes en el periodo transcurrido entre la realización del CNPV y de la EIH.

- iii) El ordenamiento de los dominios y departamentos coincide plenamente: el área rural es el dominio que presenta el mayor porcentaje de población con ingresos inferiores al valor de la línea de pobreza, y San Pedro es el departamento que registra el porcentaje más alto: 62% según ambas fuentes.¹¹

Cuadro 1
PARAGUAY: POBLACIÓN EN SITUACIÓN DE POBREZA PORCENTAJES ESTIMADOS CON INFORMACIÓN DE LA ENCUESTA INTEGRADA DE HOGARES (EIH) Y DEL CENSO NACIONAL DE POBLACIÓN Y VIVIENDAS (CNPV-MODELOS DE INGRESO)

Dominios/ Departamentos	Fuente de información utilizada				Ordena- miento
	EIH		CNPV-Modelos de ingreso		
	%	Error estándar	%	Error estándar	
Dominios					
Asunción	11.8	0.019	13.1	0.0052	4
Central urbano	20.3	0.022	25.2	0.0147	3
Resto urbano	30.1	0.016	34.1	0.0107	2
Rural	42.5	0.014	42.7	0.0078	1
Departamentos					
Asunción	11.8	0.019	13.1	0.0052	7
San Pedro	62.2	0.029	62.2	0.0162	1
Caguazú	60.3	0.026	60.6	0.0219	2
Itapúa	30.1	0.028	36.8	0.0309	4
Alto Paraná	19.5	0.025	26.3	0.0303	5
Central	16.9	0.018	20.7	0.0121	6
Resto	40.7	0.019	39.5	0.0089	3
Total	32.1	0.009	34.4	0.0042	

Ordenamiento = de mayor a menor porcentaje de población pobre.

Para el cálculo de los errores estándar se utilizó, en el caso de la EIH, la metodología incorporada en CENVAR (Census Variance Calculation System) del software IMPS y, en el caso del CNPV-Modelos, las fórmulas propuestas por Lanjouw y Lanjouw "Estimators of poverty and inequality measures using combined data sources", mimeo, Vrije Universiteit, citadas en J. Hentschel y otros (2000).

11 Con la EIH y utilizando pruebas de dominancia estocástica, Robles (1999b) ha mostrado que a nivel de dominios geográficos se cumple la clasificación presentada en el cuadro 1, independientemente del valor de la línea de pobreza y la medida de pobreza utilizada.

Con estos resultados puede responderse ahora a las siguientes preguntas: ¿en qué lugar del país se encuentran los hogares con menores niveles de ingreso o consumo? ¿Cuáles son los distritos que tienen las más altas proporciones de población en condiciones de pobreza? ¿Hay más pobres en los lugares donde los ingresos familiares son más bajos?

Las respuestas pueden deducirse con facilidad si primero se ordenan los distritos y departamentos en función del ingreso promedio de los hogares (de menor a mayor) y de la proporción de la población en condiciones de pobreza (de mayor a menor). Según puede comprobarse en los cuadros 2 y 3 este ordenamiento permite constatar lo siguiente:¹²

- i) Los hogares con niveles de ingreso más bajos se encuentran en los distritos de San Pablo, Isidoro Resquín, 3 de Febrero, Yhu y Raúl Arsenio Oviedo; los con niveles más altos, en Asunción, Ciudad del Este, Lambaré, Fernando de la Mora y San Lorenzo. Las familias de Asunción muestran un ingreso promedio que supera en siete veces el de las residentes en los diez distritos más pobres del país.
- ii) En los departamentos de San Pedro (San Pablo y Unión) y Caaguazú (3 de Febrero, San Joaquín, Yhu) se ubican los distritos con las más altas proporciones de población en condiciones de pobreza; en Central (J. Augusto Saldívar, Aregua, Itaugua, Nueva Italia) y Asunción, aquéllos con las proporciones más bajas. En los primeros, los pobres representan entre 71% y 78% de la población total, mientras que en los segundos, sólo entre 6% y 13%.
- iii) Los distritos con menores niveles de ingreso familiar no son necesariamente aquellos con mayores proporciones de población pobre. Por ejemplo, el ingreso promedio es 13% más alto en el distrito de San Pedro que en Isla Umbu (departamento de Ñeembucú), pero en este distrito el porcentaje de pobres es de 60%, mientras que en San Pedro alcanza sólo a 36%.

12 Los resultados relacionados con el consumo de los hogares no se reseñan en el presente texto, pero pueden encontrarse en los anexos E y G de Robles (1999b).

Cuadro 2
**PARAGUAY: LOS DISTRITOS CON INGRESOS FAMILIARES
MÁS ALTOS Y MÁS BAJOS**
(En guaraníes de febrero de 1998)

Departamento	Distrito	Ingreso familiar mensual estimado	Índice de ingreso ^a	Orden
San Pedro	San Pablo	408 487	100.0	1
San Pedro	Gral. Isidoro Resquín	409 151	100.2	2
Caaguazú	3 De Febrero	421 736	103.2	3
Caaguazú	Yhu	432 066	105.8	4
Caaguazú	Raúl Arsenio Oviedo	434 612	106.4	5
San Pedro	Unión	438 935	107.5	6
San Pedro	25 De Diciembre	439 251	107.5	7
Caaguazú	Carayao	445 816	109.1	8
Caaguazú	R I 3 Corrales	449 457	110.0	9
Caaguazú	La Pastora	466 493	114.2	10
Central	Villa Elisa	1 564 493	383.0	209
Alto Paraná	Hernandarias	1 583 214	387.6	210
Central	Luque	1 624 589	397.7	211
Itapúa	Encarnación	1 661 365	406.7	212
Alto Paraná	Presidente Franco	1 697 993	415.7	213
Central	San Lorenzo	1 766 756	432.5	214
Central	Fernando de La Mora	2 067 915	506.2	215
Central	Lambaré	2 080 741	509.4	216
Alto Paraná	Ciudad del Este	2 246 604	550.0	217
	Asunción	3 086 500	755.6	218

^a Ingreso familiar promedio distrital más bajo = 100.

La explicación de este resultado radica en dos importantes diferencias entre los distritos y departamentos del país: el tamaño de las familias y el porcentaje de población rural. Para un nivel de ingreso familiar dado, cuanto menor sea el número de miembros del hogar, mayor será el ingreso por persona; por otra parte, a mayor porcentaje de población rural, mayor acceso a una canasta básica de consumo (porque su costo es menor en estas áreas).

En el ejemplo anterior, San Pedro tiene, en promedio, 5.1 miembros por hogar y 84.5% de su población es rural, mientras que Isla Umbu tiene 4.0 miembros por hogar y 91.2% de población rural. El

Cuadro 3
**PARAGUAY: LOS DISTRITOS CON PORCENTAJES DE POBLACIÓN
 POBRE MÁS ALTOS Y MÁS BAJOS**

Departamento	Distrito	Proporción de la población en condiciones de pobreza ^a	Orden
San Pedro	San Pablo	0.779	1
Caaguazú	3 De Febrero	0.742	2
San Pedro	Unión	0.717	3
Caaguazú	San Joaquín	0.716	4
Caaguazú	Yhu	0.710	5
Caaguazú	R I 3 Corrales	0.710	6
Caaguazú	Simón Bolívar	0.695	7
Caaguazú	Carayao	0.694	8
San Pedro	Gral. Isidoro Resquín	0.694	9
Caaguazú	Raúl Arsenio Oviedo	0.690	10
Central	Itá	0.170	209
Central	Fernando De La Mora	0.152	210
Central	Ypacarai	0.151	211
Central	Villeta	0.141	212
Central	San Antonio	0.140	213
	Asunción	0.131	214
Central	Nueva Italia	0.125	215
Central	Itaugúa	0.116	216
Central	Aregua	0.096	217
Central	J. Augusto Saldívar	0.059	218

^a Probabilidad de que un individuo tenga un ingreso (familiar por persona) inferior al costo de una canasta básica de consumo.

costo de la canasta en las áreas urbanas de esos departamentos es 66% más alto que en las rurales.¹³

2. Comparación con el método de Necesidades Básicas Insatisfechas (NBI)

Las estimaciones de ingreso y pobreza a nivel de distritos pueden compararse con los mapas de pobreza elaborados por la DGEEC en 1995

13 Para entender estas diferencias, con el mismo procedimiento descrito aquí en relación con la incidencia de la pobreza, podría cuantificarse lo que se conoce como "intensidad" de la pobreza (la distancia entre el ingreso de los pobres y la línea de pobreza), a partir de los indicadores propuestos en Foster, Greer y Thorbecke (1984).

según el método de las NBI (STP/DGEEC/FNUAP/PNUD, 1995). En principio, ambos resultados no deberían coincidir necesariamente, debido a que responden a criterios diferentes para determinar quiénes padecen carencias o se encuentran en situación de pobreza. Sin embargo, se debería observar algún grado de correlación entre los ordenamientos distritales obtenidos con cada uno de ellos, dado que sus diferentes dimensiones generalmente se ponen de manifiesto de manera simultánea, sobre todo en los niveles más críticos de la pobreza.

El valor del coeficiente de correlación de rangos de Spearman entre los ordenamientos distritales obtenidos según la proporción de población con al menos una NBI y según la proporción de población con ingresos por debajo de la línea de pobreza muestra que el nivel de asociación alcanza a 74% (correlación positiva significativa a nivel de 1%).¹⁴ Este resultado también puede comprobarse si se estiman de manera simultánea las proporciones de pobreza a nivel distrital con ambos métodos (es decir, mediante las categorías del "método integrado" de medición de la pobreza).

En el cuadro 4 se muestra que en 81 de los 218 distritos de Paraguay predomina la "pobreza crónica", es decir, altas proporciones de población con ingresos insuficientes (más de 40% por debajo de la línea de pobreza) y NBI (más de 80% con al menos una).¹⁵ También puede observarse que en 16 distritos predomina la "pobreza inercial" (es decir, la proporción de población con NBI es mayor que la con ingresos reducidos), y en 33 distritos, la "pobreza reciente" (el porcentaje de población con ingresos reducidos supera aquél con NBI).

En el anexo H de Robles (1999b), se encuentra una lista completa de los distritos con predominio de "pobreza crónica", definidos según los cortes indicados. Resalta el hecho de que, con excepción de Asunción y Central, todos los departamentos tienen al menos un distrito en esta categoría.

14 En Perú, este resultado no alcanza a 40% (véase Robles y Reyes, 1996).

15 Definiendo otros puntos de corte se llega a agrupaciones distritales diferentes.

Cuadro 4

**PARAGUAY: POBREZA DISTRITAL SEGÚN LOS MÉTODOS
DE MEDICIÓN DE NECESIDADES BÁSICAS INSATISFECHAS (NBI)
Y LÍNEA DE POBREZA (LP)**

	Más de 40% con ingresos por debajo de la LP	Menos de 40% con ingresos por debajo de la LP	Total de distritos
Más del 80% con al menos una NBI	81 distritos (con predominio de pobreza crónica)	16 distritos (con predominio de pobreza inercial)	97
Menos de 80% con al menos una NBI	33 distritos (con predominio de pobreza reciente)	88 distritos (sin predominio de algún tipo de pobreza)	121
Total	114 distritos	104 distritos	218

III. MAPAS DE POBREZA

Con el propósito de observar la concentración geográfica de la pobreza, los resultados también se proporcionaron en mapas distritales de Paraguay, de acuerdo con sus diferentes posibilidades de representación. En principio, pudo definirse un mapa por cada combinación de: i) los diferentes criterios de que se dispone para medir la pobreza (probabilidad de ser pobre, NBI, ingreso, otros); ii) el tipo de agrupación de estos criterios (percentiles, estratos, rangos predefinidos, otros); y iii) los niveles de delimitación geográfica (departamentos, distritos, barrios, manzanas).¹⁶ En el presente estudio se consideró el nivel distrital: los indicadores de ingreso familiar per cápita, población pobre (proporción y número) y población con una NBI al menos, y los cuartiles a nivel distrital (cuatro grupos ordenados, de 54 a 55 distritos).

De los mapas 1 a 4 puede deducirse lo siguiente:

- i) La clasificación de los distritos de acuerdo con un determinado indicador de bienestar no coincide plenamente con la clasificación obtenida con otros indicadores. Por ejemplo, de los 54 distritos que se ubican en el cuartil con mayor porcentaje de población pobre, 37 corresponden al cuartil con ingreso familiar más bajo, 21 al cuartil

¹⁶ Con diferentes márgenes de error podría llegarse hasta el nivel de barrios y manzanas, pues las estimaciones se elaboraron a nivel de hogares censados.

con más población pobre y 27 al cuartil con mayor porcentaje de población con un mínimo de una NBI.

- ii) Si se consideran los cuatro indicadores utilizados, sólo cinco distritos se encuentran en el cuartil más bajo. Es decir, los únicos distritos que tienen, simultáneamente, los ingresos familiares más bajos, el mayor porcentaje y número de pobres y el mayor porcentaje de población con al menos una NBI son 3 de Febrero, San Joaquín, Yhu y Carayao en Caaguazú y General Isidoro Resquín en San Pedro. Definitivamente, se encuentran en las peores condiciones socioeconómicas.
- iii) San Pablo es el distrito con mayor porcentaje de población pobre y, a la vez, el que registra el más bajo promedio de ingreso familiar. Puerto Pinasco es el que tiene el más alto porcentaje de población con al menos una NBI. ¡Y Asunción es el distrito con el mayor número de pobres!
- iv) La selección del indicador de bienestar para elaborar el mapa es importante no sólo para facilitar la observación de la población destinataria, sino también para definir el énfasis del programa o política social que se elija. Es decir, un indicador será mejor que otros, dependiendo del objetivo que se persiga. Si, por ejemplo, lo que se desea es mejorar la distribución del ingreso en el país, el indicador seleccionado podría ser el ingreso familiar o el porcentaje de hogares pobres; pero si lo que se busca es más bien reducir la masa de pobres en un porcentaje determinado, el indicador debería ser el número de pobres. Obviamente, si lo que se pretende es lograr ambos objetivos, el mapeo tendría que incorporar más de un indicador.
- v) En este sentido, según puede inferirse de la información en que se basan los mapas, un objetivo no necesariamente coincide con el otro. Por ejemplo, si la meta fuera reducir la pobreza en un porcentaje determinado y el gasto social se focalizara en los 20 distritos con más altos porcentajes de población pobre (San Pablo, 3 de Febrero, Unión, San Joaquín, Yhu, otros) la incidencia de la pobreza en el país se reduciría del 33% actual a 28%.

Sin embargo, si las acciones se focalizaran en los 20 distritos con mayor número pobres (Asunción, San Estanislao, Caaguazú, San Lorenzo, Ciudad del Este, otros), la incidencia de la pobreza se reduciría de 33% a 20%. Obviamente, la cuestión de si con uno u otro indicador se está atendiendo a los más pobres no es clara, lo cual demuestra la necesidad de utilizar otros, que midan no sólo la magnitud del problema, sino también su intensidad.

San Pedro es el departamento que registra el mayor porcentaje de población pobre y el nivel de ingreso familiar más bajo, pero el mayor número de pobres se localiza en Caaguazú. El departamento Central, si bien está entre los con menores porcentajes de población pobre, es uno de los que tienen mayor número de pobres (cuadro 5).

Cuadro 5
PARAGUAY: INGRESOS FAMILIARES Y POBREZA POR DEPARTAMENTOS

Departamento	Población pobre				Ingreso familiar mensual estimado		Población con al menos una NBI	
	Propor- ción ^a	Orden ^b	Número	Orden ^c	(Guaraníes de feb.1998)	Orden ^d	Propor- ción	Orden ^b
San Pedro	0.622	1	220 897	3	504 692	1	0.808	5
Caaguazú	0.606	2	278 630	1	636 633	5	0.796	8
Caazapa	0.481	3	69 557	11	580 695	2	0.803	6
Concepción	0.478	4	91 578	6	739 974	8	0.833	4
Alto Paraguay	0.446	5	6 483	18	612 179	3	0.942	1
Guairá	0.407	6	71 836	10	754 669	9	0.769	11
Amambay	0.394	7	53 089	13	906 142	13	0.772	10
Canindeyú	0.385	8	55 049	12	624 838	4	0.797	7
Misiones	0.383	9	38 569	14	832 729	12	0.735	14
Itapúa	0.368	10	177 751	4	930 899	14	0.729	15
Ñeembucú	0.365	11	32 329	15	734 941	7	0.791	9
Pte. Hayes	0.362	12	29 641	16	801 080	11	0.889	2
Cordillera	0.352	13	76 684	8	775 819	10	0.741	13
Paraguarí	0.345	14	86 274	7	731 998	6	0.745	12
Boquerón	0.329	15	12 411	17	1 063 839	15	0.883	3
Alto Paraná	0.263	16	176 870	5	1 456 634	16	0.683	16
Central	0.207	17	266 878	2	1 604 246	17	0.627	17
Asunción	0.131	18	73 619	9	3 086 500	18	0.369	18

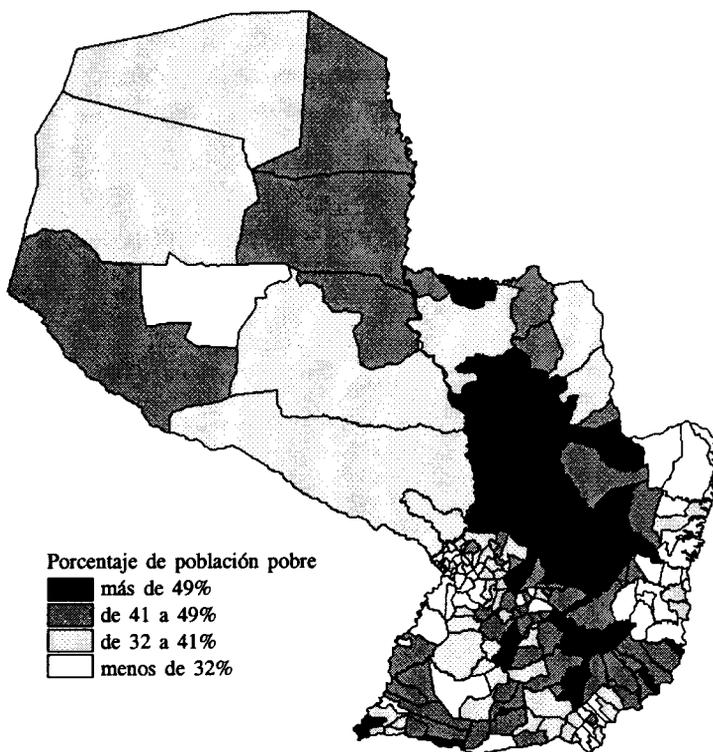
^a Probabilidad de ser pobre (ingreso por persona inferior al costo de una canasta básica de consumo.

^b De mayor a menor proporción.

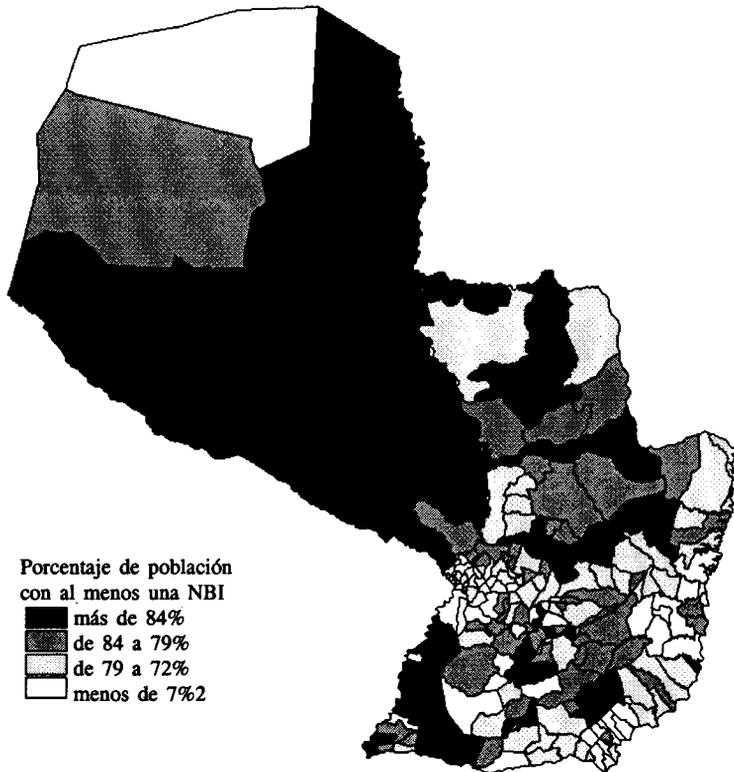
^c De mayor a menor número (a partir de la población ajustada de 5 405 474 personas).

^d De mayor a menor ingreso.

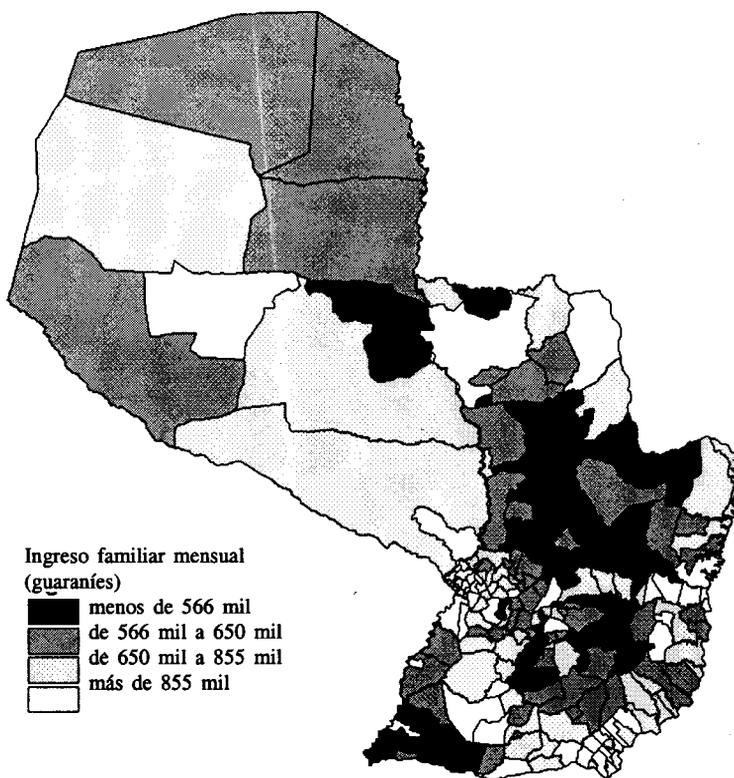
MAPA 1
PARAGUAY: PORCENTAJE DE LA POBLACIÓN
CON INGRESOS POR DEBAJO DE LA LÍNEA
DE POBREZA, SEGÚN DISTRITOS



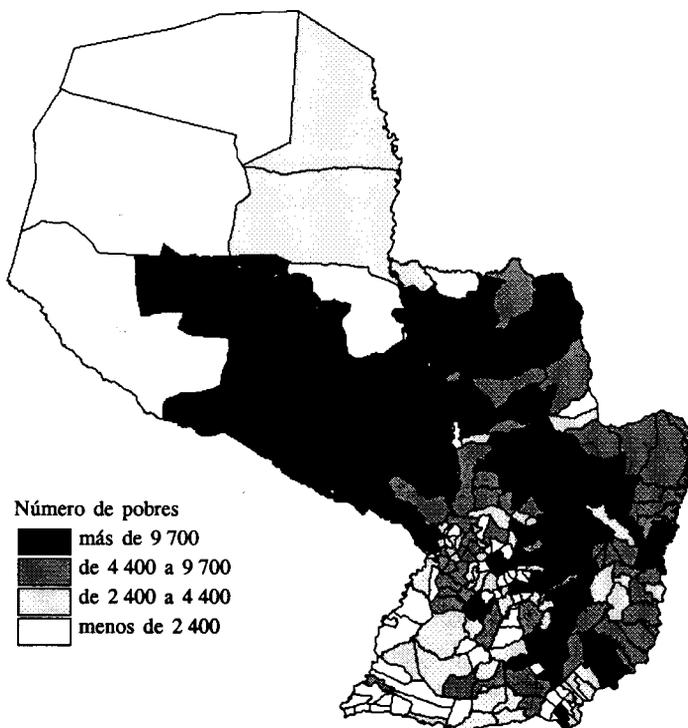
MAPA 2
PARAGUAY: PORCENTAJE DE POBLACIÓN
CON AL MENOS UNA NBI, SEGÚN DISTRITOS



MAPA 3
PARAGUAY: INGRESO FAMILIAR PROMEDIO
MENSUAL, SEGÚN DISTRITOS
(En guaraníes de febrero de 1998)



MAPA 4
PARAGUAY: NÚMERO DE HABITANTES CON INGRESOS
POR DEBAJO DE LA LÍNEA DE POBREZA, SEGÚN DISTRITOS



1. ¿Cómo asignar los recursos de un programa social a nivel de distritos?¹⁷

Si no se dispone de un indicador específico de bienestar a nivel de distritos, cabría pensar en una asignación según el tamaño de la población, sobre la base de la última información censal. Si bien es cierto que ésta es una alternativa que podría beneficiar a muchos no pobres, es posible que sirva como referente para evaluar otras asignaciones basadas en indicadores específicos de focalización.

Sin embargo, si se dispone de un indicador de pobreza o bienestar a nivel distrital, la forma más simple de asignar los recursos de un programa social es en función del tamaño de la población ponderado por el valor de dicho indicador. En este sentido, la fórmula para cada uno de los 218 distritos de Paraguay sería la siguiente:

$$\% \text{ asignado al distrito}_i = \frac{(\text{Indicador}_i * \text{Población}_i)}{\sum_{i=1}^{218} (\text{Indicador}_i * \text{Población}_i)}$$

Con esta fórmula, si el indicador fuera el porcentaje de población con al menos una NBI, al distrito *i* le correspondería una asignación proporcional al número de personas con carencias allí radicadas. Por ejemplo, a Puerto Pinasco (Presidente Hayes), que es el distrito con el porcentaje más alto de NBI (98%), y cuenta con 3 972 habitantes, se le asignaría $98.0 * 3972 / 371216700 = 0.105\%$ del presupuesto total del programa, mientras que a Asunción, el distrito con el más bajo porcentaje de NBI (36.9%), y con 562 934 habitantes, recibiría $36.9 * 562934 / 371216700 = 5.599\%$. De este modo, la asignación per cápita en el distrito de Puerto Pinasco sería 2.7 veces mayor que en Asunción.

Interesa señalar que con la fórmula descrita todos los distritos podrían beneficiarse, ya que cada uno de ellos tiene un porcentaje de población pobre o con carencias. Sin embargo, es posible aplicarla sólo a un grupo de distritos considerados "prioritarios", como generalmente se hace en el contexto de los programas de lucha contra la pobreza. Por ejemplo, podría designarse como grupo prioritario aquellos distritos pobres en términos de insuficiencia de ingresos y, a la vez, de necesidades básicas insatisfechas; otra posibilidad sería integrar los distritos con las proporciones más altas de pobreza y el mayor número de pobres. En la evaluación de los indicadores de focalización se considerarían estas dos alternativas (véase la próxima sección).

17 En esta sección se sigue el razonamiento utilizado por Schady (1998).

Para las simulaciones se supone que se cumplen las siguientes condiciones:

- i) que el beneficio por persona es el mismo para todos los distritos;
- ii) que los beneficios destinados a la población de un distrito no son aprovechado por residentes de otros distritos;
- iii) que los beneficios se reflejan en un aumento de los ingresos individuales (ingreso familiar per cápita) y, por ende, en un acercamiento a la línea de pobreza, si se es pobre, o en un alejamiento, si no se es;
- iv) que el costo de la canasta básica de consumo (línea de pobreza) se mantiene constante, antes y después de la recepción de los beneficios del programa social;
- v) que los costos administrativos del programa son similares en todos los distritos;
- vi) que los comportamientos individuales no experimentan cambios después de recibidos los beneficios,
- vii) que no se cuenta con información completa, en el sentido de que sólo permite llegar a una asignación por distritos y no focalizada en la población pobre residente en cada uno de ellos.

2. ¿Qué indicadores de focalización alternativos podrían utilizarse para asignar los recursos de un programa social?

De acuerdo con la información disponible, se evaluaron en total siete indicadores de focalización geográfica:

- i) El porcentaje de población con al menos una necesidad básica insatisfecha, publicado en Paraguay, *Atlas de Necesidades Básicas Insatisfechas* (STP/DGEEC/FNUAP/PNUD, 1995); las necesidades consideradas se refieren a calidad de vivienda, infraestructura sanitaria, acceso a la educación y capacidad de subsistencia. Actualmente, a falta de otros indicadores, en Paraguay se estaría utilizando la información proporcionada por este esquema para focalizar algunos programas sociales.¹⁸

18 En BID (1996), se indica que "En busca de hacer el mejor uso de la información disponible y minimizar el costo de la focalización, se utilizan los datos ya existentes sobre necesidades básicas insatisfechas. Así, el método adoptado para determinar si los beneficiarios son de bajos ingresos consiste en investigar si el indicador de necesidades básicas insatisfechas (NBI) en el distrito donde resida la población beneficiaria, correspondiente al tipo de inversión solicitada, es peor que el promedio nacional".

- ii) El nivel de ingreso familiar disponible, estimado sobre la base de una combinación de datos del CNPV y la EIH y los modelos econométricos correspondientes.
- iii) El porcentaje de población residente en hogares cuyos ingresos son inferiores al costo de la canasta básica de consumo (o a la línea de pobreza), estimación que se basa en datos combinados del CNPV y la EIH y los modelos econométricos correspondientes.
- iv) El número de personas que vive en hogares cuyos ingresos son inferiores a la línea de pobreza, estimado con información tanto del CNPV como de la EIH y los modelos econométricos correspondientes.
- v) Se identifican así 81 distritos con predominio de pobreza crónica (combinación de i) y iii)), es decir, aquellos que muestran, a la vez, más de un 40% de población con ingresos por debajo de la LP y más de un 80% con al menos una NBI.
- vi) También 82 distritos pertenecientes al cuartil con más altos porcentajes de población con ingresos inferiores a la LP y, simultáneamente, al cuartil con mayor número de personas con ingresos inferiores a la LP.
- vii) La población del distrito con respectos a la total.
 Junto con los elegidos aquí se podría evaluar otros indicadores y combinaciones, en la medida en que se dispusiera de información adicional a nivel de distritos.

3. ¿Cómo evaluar las bondades de los distintos indicadores de focalización geográfica?

Para evaluar los indicadores de focalización se combinó información distrital respecto de los indicadores descritos con datos sobre ingresos familiares procedentes de la EIH 1997/1998 (además de los valores de la línea de pobreza estimados por la DGEEC).

En principio, para simular asignaciones distritales se utilizó la fórmula descrita anteriormente, y luego a nivel per cápita. Es decir, para los siete indicadores: i) se estimó la proporción de los recursos de un programa que podría destinarse a cada distrito; ii) se dividió esta proporción entre la población total de cada distrito para obtener la proporción per cápita; y iii) este resultado se multiplicó por el monto total del programa para obtener la transferencia per cápita.

Luego se imputó esta transferencia (la correspondiente a cada indicador) a los hogares de cada distrito considerado en la EIH 1997/1998 (4 353 hogares pertenecientes a 164 distritos de los 218 que existen en Paraguay). Estas imputaciones, junto con las demás variables de la encuesta, configuran una base de datos representativa de todos los hogares del país, exactamente como en el caso de la base original de la EIH; por lo tanto, a partir de ella pueden evaluarse los indicadores de focalización con niveles similares de confianza estadística.

Respecto de los criterios para evaluar los indicadores de focalización, si lo que se busca es reducir los niveles de pobreza, una condición fundamental es usar de manera eficiente los recursos escasos del Estado. En el caso de los que se destinan al gasto social, esto implica lograr que los beneficios generados lleguen de la mejor manera posible a la población pobre y, a la vez, que no lleguen a la población no pobre.

Para lograrlo, los indicadores elegidos deberían evaluarse con respecto a lo que en la bibliografía sobre el tema se conoce como errores tipo I y II, la tasa de filtración y la de subcobertura (Grosh, 1994; otros); también habría que simular transferencias a los hogares en función de dichos indicadores.¹⁹ Si la población se clasifica en pobres y no pobres, así como en beneficiados y no beneficiados, el error tipo I (de exclusión), será igual al número de pobres no beneficiados con respecto a la población total; el error tipo II (de inclusión), igual al número de no pobres beneficiados en relación con la población total. La tasa de filtración será igual al número de no pobres beneficiados respecto del total de beneficiados y la tasa de subcobertura igual al número de pobres no beneficiados en relación con el total de pobres. Un indicador será mejor que otro si muestra errores de menor magnitud.

Los indicadores elegidos se evaluaron asimismo en términos de los cambios que registrarían las tasas de pobreza si se simularan transferencias de los programas a los hogares, es decir, la medida en que variaría la pobreza inmediatamente después de ejecutada una transferencia. Un indicador será mejor que otro si ejerce un mayor impacto sobre los ingresos familiares y, por lo tanto, en la pobreza.

Para esta simulación se consideraron las medidas de pobreza correspondientes a los tres primeros miembros de lo que se conoce como la familia de indicadores FGT: incidencia, brecha y severidad de la pobreza (Foster, Greer y Thorbecke, 1984). La **incidencia** se mide

19 Schady (1998) también utiliza las llamadas curvas de concentración y estimaciones de densidades no paramétricas del ingreso per cápita.

por la proporción de población residente en hogares cuyo ingreso per cápita es inferior a la línea de pobreza y, en consecuencia, denota la extensión o predominio de la pobreza. La **brecha** es el promedio de las distancias relativas entre los ingresos y la línea de pobreza (que en el caso de los no pobres es igual a cero), lo que acusa la profundidad de la pobreza. Puede expresarse también como el promedio de las distancias relativas entre los ingresos de los pobres y la línea (**intensidad** de la pobreza), multiplicado por la incidencia de la pobreza. La **severidad** de la pobreza corresponde a la brecha estimada al dar mayor peso a las distancias relativas entre los ingresos de los más pobres y la LP; por consiguiente, es una medida de la distribución del ingreso entre los pobres. En suma, estas medidas indican cuántos son los pobres, en qué medida son pobres y cuáles son las diferencias entre ellos.

En los cuadros 6, 7 y 8 y los gráficos 1 al 4 se muestran los resultados de este ejercicio. Al interpretarlos se debe prestar especial consideración al supuesto de que no se dispone de información completa, en el sentido de que sólo se llega a una asignación por distritos y no según la población pobre existente en cada uno de ellos. Los hallazgos más importantes son los siguientes:

- i) Si el gasto de un programa se focalizara según el criterio de los ingresos familiares o del número de pobres, Asunción recibiría más del doble de los recursos (26% y 28%, respectivamente) en relación con su cuota de población (10.1% del total), y más de seis veces respecto de su tasa de pobreza (4%). Sin embargo, el área rural, con estos mismos criterios, recibiría sólo la mitad de su proporción de pobreza (28%-30% en comparación con 61%). Si la focalización se hiciera sobre la base del indicador que combina los porcentajes de población con al menos una NBI e ingresos por debajo de la línea de pobreza (predominio de pobreza crónica), el área rural recibiría una proporción de los recursos mayor que su tasa de pobreza (véase el cuadro 6).
- ii) La tasa de filtración (porcentaje de beneficiados no pobres), el error de inclusión o tipo II (no pobres beneficiados con respecto a la población total) y la tasa de subcobertura (porcentaje de pobres no beneficiados) son mayores si un programa social se focaliza a partir del indicador que acusa predominio de pobreza crónica en vez de emplear el que indica mayor proporción y mayor número de pobres. Pero el error de exclusión o tipo I (pobres no beneficiados con respecto a la población total) es menor si se aplica el indicador de predominio de pobreza crónica (véase el cuadro 9). Los resultados

Cuadro 6
**PARAGUAY: DISTRIBUCIÓN DE LOS RECURSOS DE UN PROGRAMA
 SOCIAL EN FUNCIÓN DE INDICADORES ALTERNATIVOS
 DE FOCALIZACIÓN POR DISTRITOS, SEGÚN
 DOMINIOS GEOGRÁFICOS**
 (Porcentajes)

Indicador distrital	Asunción	Central urbano	Resto urbano	Área rural	Total
Necesidades básicas insatisfechas (NBI)	5.7	17.1	25.7	51.6	100.0
Ingresos familiares	25.5	24.0	23.0	27.5	100.0
Porcentaje de población pobre según LP	4.0	12.4	25.4	58.3	100.0
Número de pobres según LP	28.0	20.9	21.3	29.9	100.0
Predominio de pobreza crónica	0.0	0.0	27.9	72.1	100.0
Mayor porcentaje y mayor número de pobres	1.2	9.5	27.4	62.0	100.0
Tamaño de la población	10.5	18.5	24.7	46.3	100.0
<i>Pobreza</i>					
Distribución	3.9	11.7	23.1	61.2	100.0
Incidencia	11.8	20.3	30.1	42.5	32.1

para el resto de los indicadores son: error tipo I y tasa de filtración igual a 68%; error tipo II y tasa de subcobertura igual a 0%.

- iii) El desvío de los beneficios de un programa social hacia los no pobres y la limitada cobertura de los pobres pueden captarse mejor si se considera el destino de los montos distinguiendo entre indicadores de focalización. Al hacerlo se confirma (véase el cuadro 7) que el indicador de predominio de pobreza crónica es el que permite que los recursos lleguen más focalizadamente a la población pobre; le sigue en importancia el indicador que combina el porcentaje de población y el número de personas con ingresos inferiores a la línea de pobreza. El indicador de focalización que menos contribuye a que los recursos lleguen a los pobres es el de ingreso familiar (como ya se indicó, está sesgado por el hecho de no considerar el tamaño familiar, que es mucho mayor en los estratos bajos).

Cuadro 7
**PARAGUAY: DISTRIBUCIÓN DE LOS RECURSOS DE UN PROGRAMA
SOCIAL EN FUNCIÓN DE INDICADORES ALTERNATIVOS
DE FOCALIZACIÓN POR DISTRITOS,
SEGÚN CONDICIÓN DE POBREZA**
(Porcentajes)

Indicador distrital	No pobre	Pobre	Total
Ingresos familiares	77.4	22.6	100.0
Tamaño de la población	67.9	32.1	100.0
Número de pobres según LP	69.5	30.5	100.0
Necesidades básicas insatisfechas	64.8	35.2	100.0
Porcentaje de población pobre según LP	58.4	41.6	100.0
Mayor porcentaje y mayor número de pobres	53.6	46.4	100.0
Predominio de pobreza crónica	48.2	51.8	100.0

- iv) Una transferencia anual a los hogares de un billón de guaraníes haría que la tasa de pobreza se redujera de 32% a 23.7% si como indicador de focalización se utilizara el número de pobres (véase el cuadro 8). El uso de los indicadores de predominio de pobreza crónica e ingreso familiar disminuiría la tasa de pobreza sólo a 27%, menos que todos los demás. Como se señaló anteriormente, el indicador de número de pobres presenta el sesgo de concentrar los recursos en las zonas donde la población pobre es más densa, como Asunción (área que presenta el mayor número, aunque una de las más bajas proporciones de pobres respecto de su población total), o las áreas urbanas en general, en detrimento de poblaciones con mayor incidencia de la pobreza. Es indudable que la búsqueda de una solución de compromiso (*trade-off*) entre número de pobres y porcentaje de población pobre lleva finalmente a admitir la necesidad de evaluar indicadores combinados.
- v) El monto de los recursos disponibles para la ayuda social es un elemento importante que debe tenerse en consideración al seleccionar el indicador de focalización si el objetivo es realmente ejercer el mayor impacto posible sobre las tasas de pobreza (véanse los gráficos 1 a 4). Por ejemplo, cuando los montos que se transferirán son elevados, el indicador de proporción de población con ingresos inferiores a la línea de pobreza es mucho más eficiente

Cuadro 8
**PARAGUAY: VARIACIONES EN LA INCIDENCIA DE LA POBREZA
 COMO RESULTADO DE LA TRANSFERENCIA ANUAL
 A LOS HOGARES DE MIL MILLONES DE GUARANÍES,
 SEGÚN INDICADORES DE BIENESTAR Y ÁREAS
 GEOGRÁFICAS DE RESIDENCIA**

Indicador distrital	Urbano	Rural	Total
Necesidades básicas insatisfechas (NBI)	20.0	32.7	25.9
Ingresos familiares	17.2	38.1	26.9
Porcentaje de población pobre según LP	20.0	30.4	24.8
Número de pobres según LP	16.0	32.7	23.7
Predominio de pobreza crónica	21.4	33.4	26.9
Mayor porcentaje y mayor número de pobres	20.0	32.3	25.7
Tamaño de la población	19.4	33.7	26.0
Situación inicial de pobreza	23.1	42.5	32.1

- para reducir la tasa de pobreza; pero si los montos son más reducidos, conviene emplear el indicador de número de pobres.
- vi) Con respecto a las demás medidas de pobreza (intensidad, brecha y severidad), se observa que para los programas sociales que buscan reducir la pobreza, el peor indicador de focalización es el de predominio de pobreza crónica, aunque sólo a partir de los 800 000 millones de guaraníes; a su vez, después del indicador de proporción de población con ingresos inferiores a la línea de pobreza, el que tiene más impacto es el de NBI, aunque también a partir de mil millones de guaraníes en el caso de las medidas de intensidad y brecha de la pobreza, y a partir de 700 000 millones de guaraníes en relación con la severidad de la pobreza.

Cuadro 9
**PARAGUAY: EVALUACIÓN DE UN PROGRAMA SOCIAL
 SEGÚN EL INDICADOR DE FOCALIZACIÓN UTILIZADO**
(Porcentajes)

Indicador a nivel de distrito	Error tipo		Filtración	Subcobertura
	I	II		
Ingresos familiares	67.9	0.0	67.9	0.0
Tamaño de la población	67.9	0.0	67.9	0.0
Número de pobres según LP	67.9	0.0	67.9	0.0
Necesidades básicas insatisfechas (NBI)	67.9	0.0	67.9	0.0
Porcentaje de población pobre según LP	67.9	0.0	67.9	0.0
Mayores porcentaje y número de pobres	48.5	8.4	6.3	26.3
Predominio de pobreza crónica	7.8	22.0	43.6	68.6

IV. CONCLUSIONES

- Es factible obtener estimaciones coherentes y confiables de indicadores alternativos al de NBI a escala de distritos mediante la combinación de fuentes de información con características diferentes (en términos de niveles de inferencia y temas abordados). Ha sido suficiente contar con un conjunto mínimo de variables comunes y un marco analítico adecuado que permitiera verificar la confiabilidad de las estimaciones.
- Al combinar fuentes de información es posible interceptar resultados de mediciones de pobreza basadas en criterios distintos y, por lo tanto, abordar este problema en sus múltiples dimensiones y con toda la heterogeneidad que presenta en el país. En este sentido, se ha constatado que existen coincidencias importantes (o complementariedades) en el ordenamiento de los distritos del país, independientemente de los métodos de medición de la pobreza que se apliquen.
- La importancia de estas constataciones se vuelve evidente cuando el objetivo es una mejor focalización del gasto social. Como se sabe, dependiendo del criterio utilizado para medir la pobreza, el énfasis de las políticas podría apuntar fundamentalmente a mejorar ya sea las condiciones de empleo y el consumo de los hogares o

bien el acceso a los servicios públicos. Según se ha visto, en algunos ámbitos predomina una pobreza de tipo coyuntural y en otros, de tipo estructural.

- Las perspectivas respecto de la implementación de un conjunto de reformas económicas en el país marcan un momento propicio para buscar formas y procedimientos que permitan mejorar la inversión social mediante un mayor énfasis en la eficiente asignación de los recursos. Para ello es fundamental reducir los problemas de exclusión (no llegar a los pobres) e inclusión (llegar a los no pobres), tarea en la que quizás el elemento más importante sea la información desagregada geográficamente, que permite una focalización más precisa en los grupos que necesitan ser atendidos.

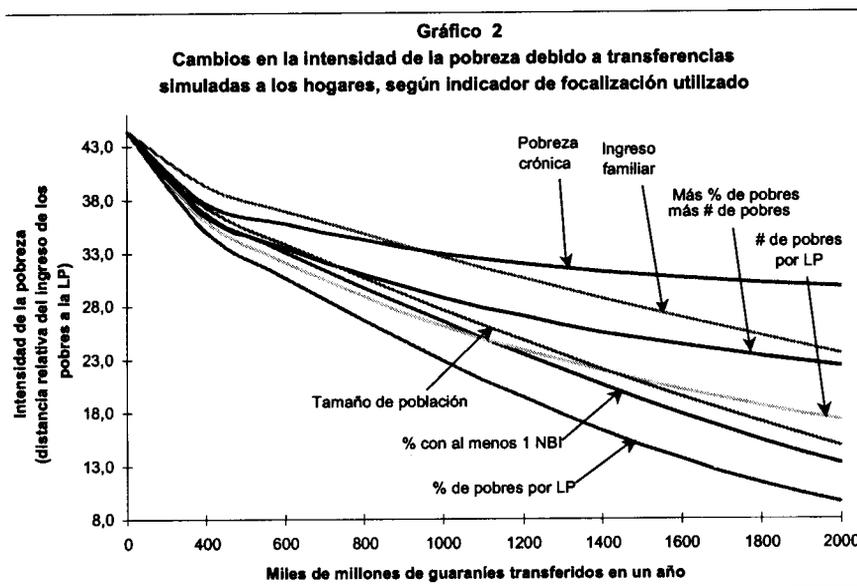
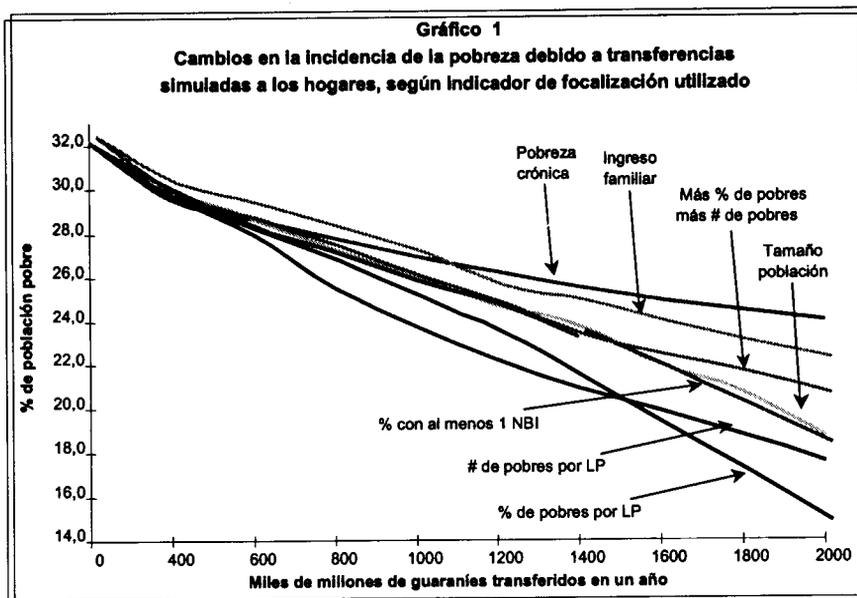


Gráfico 3

Cambios en la brecha de la pobreza debido a transferencias simuladas a los hogares, según indicador de focalización utilizado

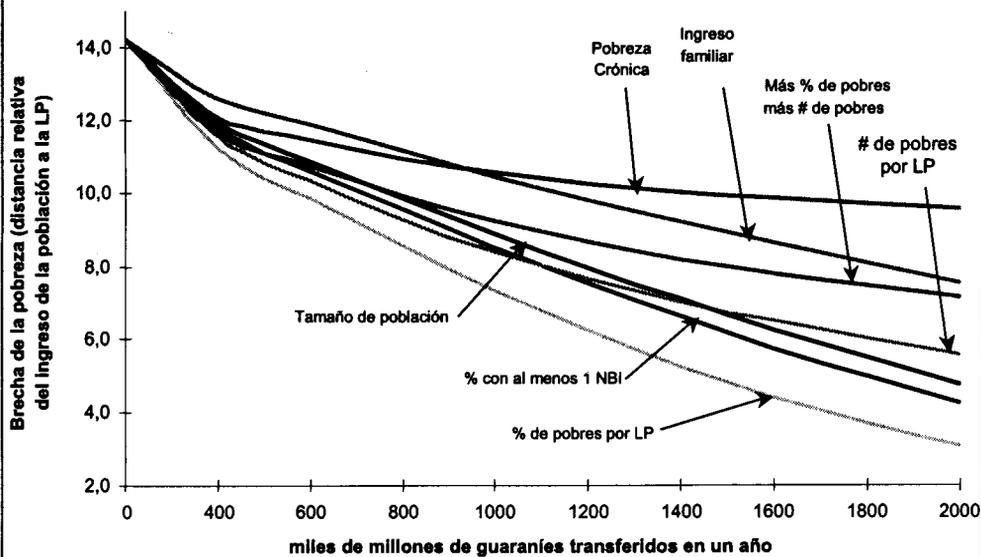
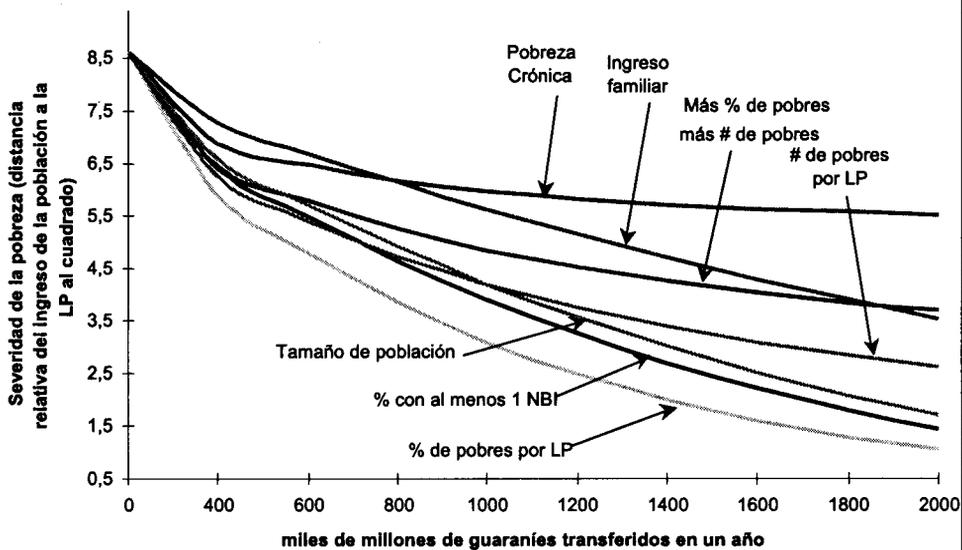


Gráfico 4

Cambios en la severidad de la pobreza debido a transferencias simuladas a los hogares, según indicador de focalización utilizado



BIBLIOGRAFÍA

- BID (Banco Interamericano de Desarrollo) (1996), "Paraguay. Propuesta de préstamo y cooperación técnica no reembolsable para un Programa de Inversiones Sociales", Operación N° 929-OC-PR/ ATN/SF-5228-PR, Washington, D.C., 22 de mayo.
- Deichmann, U. (1999), "Geographic Aspects of Inequality and Poverty", Washington, D.C., Banco Mundial, inédito.
- Demombynes, G. (1999), "A Manual for the Poverty and Inequality Mapper Module", Banco Mundial, septiembre, inédito.
- Foster, James, Joel Greer y Erik Thorbecke (1984), "A class of descomposable poverty measures", *Econométrica*, vol. 22.
- Glewwe, P. y J. van der Gaag (1988), "Confronting poverty in developing countries. Definitions, information, and policies", LSMS WP N° 48, Washington, D.C., Banco Mundial.
- Grosh, M. (1994), "Administering Target Social Programs in Latin América. From Platitudes to Practice", Washington, D.C., Banco Mundial, inédito.
- Hentschel, J. y otros (2000), "Combining Census and Survey Data to Study Spatial Dimensions of Poverty: A Case Study of Ecuador", *The World Bank Economic Review*, N° 1, vol. 14, Washington, D.C., Banco Mundial, enero.
- Robles, M. (1999a), "Canasta básica de alimentos y línea de pobreza. EIH 1997/98", Asunción, Dirección General de Estadística, Encuestas y Censos (DGEEC)/Programa para el Mejoramiento de las Encuestas y la Medición de las Condiciones de Vida en América Latina y el Caribe (MECOVI), febrero.
- (1999b), "Pobreza y distribución del ingreso en Paraguay, 1997/1998", Asunción, Dirección General de Estadística, Encuestas y Censos (DGEEC)/Programa para el Mejoramiento de las Encuestas y la Medición de las Condiciones de Vida en América Latina y el Caribe (MECOVI), julio.
- Robles, M. y J. Reyes (1996), *Propuesta metodológica para determinar el ingreso y la proporción de hogares pobres a nivel provincial y distrital*, Lima, Banco Interamericano de Desarrollo (BID)/ Centro Latinoamericano de Demografía (CELADE)/Instituto Nacional de Estadística e Informática (INEI).
- Schady, N. (1998), *Picking the Poor: Indicators for Geographic Targeting in Peru*, Princeton, Princeton University Press/Banco Mundial, octubre.

- STP/DGEEC (Secretaría Técnica de Planificación del Desarrollo Económico y Social/Dirección General de Estadística, Encuestas y Censos) (1998), *Cuestionario de la Encuesta Integrada de Hogares 1997/1998*, Asunción.
- ____ (1992), *Cuestionario del Censo Nacional de Población y Viviendas de 1992*, Asunción.
- ____ (1991), *Necesidades Básicas Insatisfechas (NBI)*, Asunción, abril.
- STP/DGEEC/FNUAP/PNUD (Secretaría Técnica de Planificación del Desarrollo Económico y Social/ Dirección General de Estadística, Encuestas y Censos/Fondo de Población de las Naciones Unidas/ Programa de las Naciones Unidas para el Desarrollo) (1995), *Paraguay. Atlas de Necesidades Básicas Insatisfechas (NBI)*, Asunción.