

EFFECTOS CONTEXTUALES Y FECUNDIDAD MARITAL: UN MODELO DE NIVELES MÚLTIPLES DE LA PARIDEZ EN LA REGIÓN DEL BIOBÍO, CHILE

José Manuel Merino Escobar
Universidad de Concepción

RESUMEN

Éste es un estudio de niveles múltiples, en que se aplicó un nuevo método de modelización estadística para determinar los efectos contextuales que influyen en la conducta reproductiva de las mujeres residentes en la Octava Región del país (Región del Biobío). Los estudios de niveles múltiples o de efectos contextuales son una línea de investigación de frontera en el campo de las ciencias sociales, que se ha estado consolidando a partir de los años ochenta en áreas como la educación, la salud, la fecundidad, las migraciones, la economía y, en general, en temas cuyos datos tienen una estructura jerárquica. Una jerarquía debe aquí ser entendida como una organización formada por unidades agrupadas en diferentes niveles: las mujeres fértiles son las unidades de nivel 1 agrupadas dentro de sus distritos rurales de residencia, los cuales pasan a ser las unidades de nivel 2, que a su vez están incluidas territorialmente en las comunas, que constituyen el nivel 3 de esta jerarquía.

En este estudio se procuró relacionar información sobre la conducta reproductiva de mujeres fértiles (nivel 1 o micro) con propiedades macroestructurales presentes tanto en los distritos rurales en que aquéllas residen (nivel 2) como en las comunas de la Octava Región (nivel 3) en que esos distritos están incluidos. El problema específico investigado fue en qué medida una variable dependiente de nivel individual como el número total de hijos obedece a diferencias

entre distritos, entre comunas o entre ambos, esto es, a características inherentes a la localidad y a la comuna en que residen, y, por otra parte, cuánto era atribuible, dentro de cada localidad, a lo que puede ser denominado estrictamente variación individual. El método que se utilizó es la modelización de niveles múltiples de regresión de Poisson, mediante la aplicación de análisis jerárquico con efectos aleatorios.

(FECUNDIDAD)
(MODELOS)

(COMPORTAMIENTO REPRODUCTIVO)

**CONTEXTUAL EFFECTS AND MARITAL FERTILITY:
A MULTILEVEL MODEL OF PARITY IN THE
BIOBÍO REGION, CHILE**

ABSTRACT

This is a multilevel analysis that apply a new statistical modelling method to determine contextual effects acting on the reproductive behaviour of women living in the Eighth Region of Chile, the Biobío Region. Multilevel analyses or contextual effects studies are a recent research paradigm within the social sciences which has been of increasing importance since the 1980s in areas such as education, health, fertility, migration, economics and, generally speaking, in areas having a hierarchical structure. In this context, a hierarchy is a structure consisting of units ranked at different levels: for example, fertile women are level 1 units, the rural districts where they live are the level 2 units, and these in turn are grouped geographically into *comunas*, the level 3 units in this hierarchical structure.

The purpose of this study was to link data on the reproductive behaviour of fertile women (level 1 or micro-level) with the macro-structural properties of both the rural districts where they live (level 2) and of the *comunas* in the Eighth Region where those districts are located (level 3). The specific research topic was to determine the extent to which a dependent variable at the individual level, such as the total number of children ever born per woman, is accounted for by differences between districts and/or between *comunas*, or in other words, by inherent characteristics of the district or *comuna* of residence; and also to determine how much was attributable, within each district, to individual variations *stricto sensu*. The method used was one of multilevel

modelling using Poisson regression, by means of a statistical procedure known as hierarchical analysis with random effects.

(FERTILITY)
(MODELS)

(REPRODUCTIVE BEHAVIOUR)

INTRODUCCIÓN

Una de las fronteras actuales más promisorias en la investigación sobre los factores determinantes de la fecundidad es el análisis de las influencias contextuales. Este tipo de estudios, también denominado análisis de niveles múltiples (*multilevel analysis*), por abarcar simultáneamente más de una unidad de análisis, se ha convertido en las últimas dos décadas en una de las más estimulantes áreas de desarrollo de la investigación en ciencias sociales (Pullum, 1991b). Sin embargo, las investigaciones que han utilizado este método han encontrado problemas para identificar claros efectos comunitarios o contextuales que operen sobre la conducta reproductiva (Casterline, 1985; Bilsborrow y Guilkey, 1987; Billy y otros, 1989). Consideramos que la causa fundamental de este problema es de orden conceptual. En la opinión de este autor, los primeros enfoques aplicados al análisis de efectos contextuales (Boyd y Iversen, 1979; Nizamuddin, 1979; Tsui, 1985; Chayovan, 1982; Mason, Wong y Entwisle, 1983) estuvieron basados en esquemas teóricos incorrectos, por lo que generaron propuestas técnicas inapropiadas, que condujeron a resultados no significativos.

Esta investigación hace particular hincapié en una nueva perspectiva conceptual en el área de las influencias contextuales que operan sobre los determinantes próximos de la fecundidad y su propuesta consiste fundamentalmente en explicar la variación de un fenómeno mediante la partición de sus causas, entre aquellas de naturaleza macro o estructural y aquellas de naturaleza micro o individual. Las variables macro se utilizan sólo para explicar la variación del fenómeno entre grupos (*between-groups variation*) y nunca para explicar diferencias o variaciones individuales porque las características macro son comunes a todos los miembros del grupo. Las dimensiones micro o variables de nivel individual son las que deben ser utilizadas para explicar las diferencias individuales o las variaciones intragrupalas del fenómeno (*within-groups variation*). Estos dos niveles de organización de los datos

permiten también establecer vinculaciones entre ellos: algunos predictores micro pueden interactuar con algunos predictores macro o ser condicionados por éstos en sus efectos sobre la variable dependiente de nivel individual. Directamente asociada a este esquema teórico conceptual se encuentra una metodología estadística especialmente diseñada para medir los diferentes tipos de efectos contextuales, procedimiento que se denomina análisis jerárquico de covarianza con efectos aleatorios. Éste es el nuevo procedimiento para el análisis de niveles múltiples que se utilizó en el presente estudio (Pullum, 1989a, 1989b, 1991a, 1991b; Merino, 1993; Pullum y Merino, 1994).

En este proyecto se intentó específicamente medir los efectos de las variables de nivel comunitario sobre la paridez marital de las mujeres rurales en edad fértil (entre 15 y 49 años) de la Octava Región de Chile. La variable dependiente del estudio fue el número de hijos de esas mujeres. A partir de una base de datos de nivel comunitario, se seleccionó un conjunto de predictores significativos para modelar la variación intergrupala de la fecundidad marital. Esta sección constituyó el área macro de la presente investigación. Posteriormente, a partir de la información de nivel individual se seleccionaron los predictores más importantes para explicar la variación intragrupal de la fecundidad de estas mujeres. Esta parte constituye el componente micro del estudio. La conjunción de ambos enfoques, es decir, del nivel macro o análisis intergrupala y del nivel micro o análisis intragrupal, dio lugar a la investigación contextual o de niveles múltiples que se había definido como el diseño adecuado para el proyecto.

I. EFECTOS CONTEXTUALES Y CONDUCTA DEMOGRÁFICA

A pesar de dos décadas de considerable desarrollo de la investigación en ciencias sociales, el análisis de niveles múltiples aún se caracteriza por una sorprendente ausencia de hallazgos significativos. Como se indicó anteriormente, la causa primaria de esa deficiencia parece residir en una conceptualización errónea de los efectos contextuales. En esta sección se sintetizarán, comentarán y analizarán los primeros intentos de modelización que se encuentran en la literatura especializada actual, para tratar de mostrar con ello las razones de su fracaso.

En la literatura internacional se observan dos tipos de estrategias globales para desarrollar la investigación de los efectos contextuales. El enfoque más común ha sido aquel que implica la inclusión simultánea, en el mismo modelo, de las variables contextuales –que por hipótesis afectan la variable dependiente de nivel individual– y de los predictores individuales. Este modelo uniecuacional supone que la importancia del contexto social quedará expresada por el nivel de significación que alcanzan los coeficientes de cada una de las variables macro utilizadas (Blalock, 1983, 1984). Esta incorporación simultánea de ambos niveles de organización de los datos en el mismo modelo dio lugar a que diversos autores hicieran observaciones críticas, basadas en perspectivas sustantivas y de modelización estadística (Bilsborrow y Guilkey, 1987; Goldstein, 1987, 1995; Pullum, 1989a).

El principal problema es que este modelo no distingue áreas de predicción separadas para las variables de nivel contextual y para las de nivel individual. Por definición, estas variables están orientadas a captar diferentes niveles de organización de los datos. Por lo tanto, como variables, tienen una estructura de error completamente diferente y, por ende, no deberían ser consideradas simultáneamente en el mismo modelo causal. En toda la tradición del análisis contextual, sólo Mason, Entwisle y Wong (1983); Pullum (1989a, 1989b); Hirschman y Guest (1990), y Goldstein (1987, 1995) han reconocido este importante hecho. Este problema conceptual se ha traducido en problemas de estimación estadística, tales como la presencia de efectos de diseño en las estimaciones de los coeficientes (Pullum, 1991a, 1991b), y, lo que es aún peor, en los resultados contradictorios y de escasa significación que arrojó la investigación de efectos contextuales durante décadas (para una síntesis de los hallazgos de la década de 1970, véase Casterline (1985), y para los correspondientes a los años ochenta, véase Bilsborrow y Guilkey (1987)).

Un segundo tipo de estrategia para enfrentar la investigación de niveles múltiples o de efectos contextuales en conductas relacionadas con la fecundidad es la que han promovido, desde comienzos de los años ochenta, Mason, Entwisle y Wong (1983). Este método se caracteriza por definir los efectos contextuales por medio de dos componentes, uno de nivel macro y otro de nivel micro. El procedimiento consiste básicamente en utilizar la ecuación micro para determinar los predictores de la variable dependiente de nivel individual. La ecuación macro de este modelo se utiliza para estudiar la variabilidad de los efectos de los predictores micro en diferentes asentamientos

(Entwisle, Mason y Hermalin, 1986). En el componente macro, los coeficientes micro dependen estocásticamente de las variables macro, esto es, se postula que la variable agregada afecta las pendientes de los coeficientes que relacionan los predictores micro con la variable dependiente. En sentido estricto, esta estrategia se caracteriza por postular que el agregado y los predictores individuales ejercen efectos interactivos sobre la conducta demográfica estudiada (Bilsborrow y Guilkey, 1987; Billy y otros, 1989).

Aun cuando esta estrategia representa un claro avance con respecto al primer tipo de enfoque, porque, por ejemplo, separa los niveles macro y micro y relaciona estos niveles con la variación intergrupala e intragrupal, presenta varias dificultades en la interpretación de los parámetros y en la comprensión de los efectos contextuales. El principal problema de este modelo es que las variables comunitarias o grupales sólo pueden alterar los efectos de las variables individuales, en vez de ejercer efectos totales propios. No existe, entonces, una medición directa de los efectos contextuales.

En cambio, en la presente investigación se utilizó un enfoque conceptual y una modelización estadística más reciente de los efectos contextuales que influyen sobre la fecundidad. Nos referimos al análisis de niveles múltiples que toma en cuenta la variación intergrupala e intragrupal (Pullum, 1989a), que ha generado a la vez un procedimiento correspondiente de modelización estadística, denominado análisis jerárquico de covarianza con efectos aleatorios (Pullum, 1989b).

La idea central de este nuevo enfoque consiste básicamente en dividir las influencias que se ejercen sobre una variable dependiente individual o de nivel micro (Y) en influencias de nivel micro (variables de tipo X) y en influencias de nivel macro (variables de tipo Z). En este primer momento, la información es separada en dos niveles: uno para agregados (como grupos, comunidades, localidades, distritos, comunas y regiones) y otro para individuos. El paso siguiente consiste en identificar, independientemente, distintos predictores en cada uno de los niveles, a fin de explicar la mayor cantidad de variación posible. Una ecuación micro (o análisis intragrupal) se construye después para explicar la variación individual dentro del agregado, y paralelamente se desarrolla una ecuación macro para explicar la variación intergrupala (Pullum, 1989a, 1989b, 1991a). Por lo tanto, la variación total de la variable dependiente Y es desagregada en variación intergrupala (nivel macro) e intragrupal (nivel micro). Pero esto no significa que se trate de análisis paralelos a diferentes niveles de agregación de los datos, sino

más bien de un modelo de niveles múltiples que divide la variación de la variable dependiente en partes explicables a nivel individual y en partes explicables en cada nivel agregado utilizado. La suma algebraica de los totales de variación explicables en cada nivel de agregación entrega el valor total de la variación de la variable dependiente a nivel individual. Por consiguiente, los coeficientes de aquellas dos ecuaciones o componentes del modelo pueden ser interpretados sea como efectos macro, sea como efectos micro, o como interacciones.

Lo fundamental en este esquema es el enfoque conceptual. El procedimiento está basado en el carácter jerárquico de los datos, esto es, en la inclusión de los individuos dentro de agregados tales como comunidades, barrios, distritos, comunas o regiones. El punto más destacado es que una característica de nivel macro –por ejemplo, una variable de nivel comunitario– puede explicar sólo una variación entre comunidades y nunca una variación individual. De acuerdo con Pullum (1991b), la presencia o ausencia de una farmacia, por ejemplo, es una constante para cada persona dentro de una comunidad, y no puede explicar por qué una mujer usa anticonceptivos mientras otra de la misma comunidad no los usa. Es decir, las variables de nivel comunitario entrarán como efectos principales sólo en las ecuaciones de nivel macro. En forma complementaria, las variables micro o de nivel individual sólo entrarán como efectos principales en la ecuación micro. Las propiedades de los contextos sociales permitirán explicar la variación entre contextos y no las diferencias individuales. Ésta es la distinción conceptual básica del enfoque adoptado en este proyecto con respecto a otros.

II. MÉTODOS

1. Las unidades de análisis y el área de investigación

El ámbito de esta investigación está constituido por los sectores rurales de la Octava Región de Chile. Esta región está formada por cuatro provincias –Concepción, Ñuble, Biobío y Arauco– que reúnen 49 comunas, en cuyas áreas rurales viven cerca de 400 000 personas, casi la mitad de las cuales son mujeres y aproximadamente un 50% de éstas está en edades reproductivas.

Las unidades de análisis en este proyecto fueron las mujeres casadas de entre 15 y 49 años residentes en el sector rural de la mencionada región, las cuales, según el Censo Nacional de Población y

Vivienda de 1992, sumaban un total de 55 905 personas. Éstas constituyeron el componente micro o individual, que fue denominado nivel 1 en el proyecto. Las comunidades o localidades en que residen estas mujeres, representadas por los 348 distritos rurales consignados en el mismo censo constituyeron el nivel 2 del estudio. Finalmente, las 49 comunas en que están ubicados los distritos rurales constituyeron el nivel 3. Estas localidades y comunas forman el componente macro del proyecto.

2. Datos y variables dependientes

El estudio se basó en datos de nivel micro y macro provenientes de la boleta censal utilizada en el mencionado recuento. Los datos de nivel macro fueron estructurados a partir de la agregación de las características demográficas básicas y de las características socioeconómicas del hogar y de la vivienda recogidas en el censo. Los datos de nivel micro se obtuvieron a partir de la información individual de las mujeres mayores de 14 años. De ese modo, la información estadística del proyecto corresponde a la combinación de esas dos bases de datos.

A nivel macro, en el nivel 2 del estudio, la variable dependiente fue el promedio distrital de paridez marital, mientras que en el nivel 3, la variable dependiente fue el promedio comunal de dicha paridez. A nivel micro, la variable respuesta correspondiente fue el número total de hijos nacidos vivos de las mujeres casadas en edad fértil residentes en los distritos rurales.

3. La modelización estadística

La metodología estadística utilizada para modelar el número total de hijos declarados es el procedimiento de niveles múltiples desarrollado por Pullum en los últimos cinco años, y que aplicó empíricamente para estudiar la temporalidad de los primeros nacimientos en los Estados Unidos (Pullum, 1989b), así como las relaciones existentes entre la disponibilidad de servicios de planificación familiar y el uso de anticonceptivos en Guatemala (Pullum, 1991a). Merino (1993) hizo una tercera aplicación de este método de estimación en los sectores rurales de Colombia.¹ En lo que sigue, el índice i denota los individuos

¹ En Pullum y Merino (1994) puede encontrarse una descripción más completa del modelo en lo referente a tres niveles de agregación, que es la modalidad utilizada en el presente estudio.

que residen en los distritos j , que a su vez están incluidos en comunas denotadas k . Las variables genéricas se definen de la siguiente manera:

Y_{ijk} es la variable dependiente, evaluada para la persona i en el distrito j y la comuna k

X_{ijk} es una variable explicativa de nivel individual

U_{jk} es una variable explicativa, característica del distrito j en la comuna k

V_k es una variable explicativa, característica de las comunas k

El procedimiento usual implicaría combinar toda esta información en un modelo de nivel individual con el siguiente formato:

$$Y_{ijk} = f(X_{ijk}; U_{jk}; V_k)$$

donde f es una función lineal de X_{ijk} , U_{jk} y V_k .

Pullum, por el contrario, separa el rol de estas variables según los diferentes niveles de agregación. Define medias y desviaciones de las variables de la siguiente manera:

$Y_{.jk}$ es la media de Y_{ijk} en el distrito j y la comuna k

$Y_{.k}$ es la media de Y_{ijk} en la comuna k

$X_{.jk}$ es la media de X_{ijk} en el distrito j y la comuna k

$X_{.k}$ es la media de X_{ijk} en la comuna k

$U_{.k}$ es la media de U_{jk} en la comuna k

$$Y_{ijk'} = Y_{ijk} - Y_{.jk}$$

$$Y_{.jk''} = Y_{.jk} - Y_{.k}$$

$$X_{ijk'} = X_{ijk} - X_{.jk}$$

$$X_{.jk''} = X_{.jk} - X_{.k}$$

$$U_{jk''} = U_{jk} - U_{.k}$$

El uso de desviaciones de las variables respecto de su promedio distrital o comunal, según corresponda, se denomina centramiento (Boyd y Iversen, 1979).

a) *División de la variación*

Para estimar la existencia de efectos contextuales, primero se calcula la variación total original de la variable dependiente, es decir, el número total de hijos de las mujeres casadas rurales, en ausencia de cualquier predictor. Esta variación total de la variable dependiente es denominada desviación inicial (Pullum, 1991a). Luego, esa variación debe ser dividida por el número de partes correspondientes a los niveles de análisis que son utilizados en el estudio. En este caso, los distritos rurales, las comunas y las mujeres individuales fueron tomados como unidades de análisis, por lo cual la división debería separar la variación en tres niveles: i) variación dentro de los distritos; ii) variación entre los distritos pero dentro de las comunas, y iii) variación entre las comunas. La partición de la variación de la variable dependiente Y_{ijk} , correspondiente a la persona i en el distrito j y en la comuna k , debería ser expresada como sigue:

$$\Sigma(Y_{ijk}-Y_{...})^2 = \Sigma(Y_{ijk}-Y_{.jk})^2 + \Sigma(Y_{.jk}-Y_{..k})^2 + \Sigma(Y_{..k}-Y_{...})^2$$

La variación total de la variable dependiente $\Sigma(Y_{ijk}-Y_{...})^2$ es dividida en variaciones intradistritales $\Sigma(Y_{ijk}-Y_{.jk})^2$, variaciones interdistritales pero intracomunales $\Sigma(Y_{.jk}-Y_{..k})^2$ y, finalmente, en variaciones intercomunales $\Sigma(Y_{..k}-Y_{...})^2$. Se aprecia claramente en esta ecuación que cada nivel de agregación de los datos es una parte específica del mismo modelo de niveles múltiples, que está formado por la adición de cada uno de los niveles que lo componen.

b) *El modelo macro o análisis entre distritos y entre comunas*

El análisis de nivel macro de este estudio consiste básicamente en determinar cuáles predictores pueden ser modelados como determinantes significativos de la variable dependiente Y en el nivel macro correspondiente. El modelo incorpora, en cada nivel macro, sólo predictores de ese nivel. Los predictores macro en este estudio fueron designados como los bloques de variables X (agregados de variables censales de nivel individual, del hogar y de la vivienda), los bloques de variables U (agregados de variables censales al nivel distrital) y los bloques de variables V (agregados de variables censales de nivel comunal).

Dado que la variable dependiente a explicar corresponde a las frecuencias de un evento se utilizó el modelo de regresión de Poisson, por estimárselo el procedimiento estadístico más apropiado para este tipo de

casos. Existen pruebas suficientes en distintos estudios demográficos que avalan la utilización de la regresión de Poisson para modelar la variable "número total de hijos" (véanse Pullum, 1995; Land, McCall y Nagin, 1995; Gardner, Mulvey y Shaw, 1995; Espenshade y Ye, 1994).

En la regresión de Poisson se modela el logaritmo natural de la media de una variable aleatoria con distribución de Poisson como una función lineal de variables predictoras o independientes (Liao, 1994; Agresti, 1996). Estos modelos, que utilizan métodos de máxima verosimilitud para los ajustes, se revelan mucho más apropiados que los métodos ordinarios de regresión mínimo cuadrática para el análisis de la fecundidad de las mujeres, por el simple hecho de que el número de niños es siempre un número entero no negativo. En el caso de mujeres individuales, el número de hijos es pequeño y frecuentemente cero, especialmente cuando se considera un lapso corto de tiempo de exposición al riesgo de concebir. Tanto en la regresión ordinaria como en la modelización de Poisson los valores esperados pueden ser números no enteros, porque estos valores son promedios de fecundidad, pero la regresión ordinaria puede arrojar, para algunas combinaciones de predictores, valores negativos, lo que arroja una estimación obviamente incorrecta e inaceptable de la fecundidad. La distribución de Poisson ofrece otra ventaja por sobre el método de regresión mínimo cuadrática ordinaria, a saber, que tiene una varianza igual a su media aritmética. Es muy probable que la varianza del número de niños tienda a ser más grande en grupos homogéneos que tienen una media grande, que en grupos del mismo tipo pero con medias más pequeñas. Aunque las distribuciones reales no tienen usualmente igualdad de medias y varianzas, este supuesto es mucho más verosímil que el supuesto de la regresión mínimo cuadrática, que postula que la varianza del error es constante (Pullum, 1995).

Si μ denota el valor esperado para la variable dependiente Y con distribución de Poisson, y X expresa una variable explicativa, entonces el modelo de regresión de Poisson tiene la siguiente formulación exponencial:

$\mu = \exp(\sum \beta_k X_{ik})$, lo que se expresa en el modelo loglineal:

$$\log \mu = \sum \beta_k X_{ik}$$

En el presente estudio de niveles múltiples, el modelo de regresión de Poisson aplicado al componente macro adoptó las siguientes especificaciones en el nivel 2 o distrital y en el nivel 3 o comunal:

i) *Regresión de Poisson interdistrital pero intracomunal*

$$\log Y_{jk} = \beta_1 X_{jk} + \beta_2 U_{jk} + u$$

donde $\log Y_{jk}$ es el promedio distrital de hijos nacidos vivos que declararon las mujeres rurales en el censo de 1992, en escala logarítmica; β_1 es el coeficiente que estima el efecto sobre la fecundidad interdistrital de las variables de nivel individual agregadas a ese nivel (X_{jk}); β_2 es el término que expresa el coeficiente de las variables "integralmente" distritales (U_{jk}). Finalmente, u_{jk} es el término que agrupa los residuos a nivel de distrito, esto es, la variación individual de cada distrito con respecto al promedio ajustado por la ecuación de regresión.

ii) *Regresión de Poisson intercomunal*

La ecuación para la estimación se simboliza como:

$$\log Y_{..k} = \beta_1 X_{..k} + \beta_2 U_{.k} + \beta_3 V_k + v_k$$

Aquí, $\log Y_{..k}$ es el promedio comunal de niños de las mujeres rurales de la región en 1992; β_1 es el coeficiente que mide el efecto de las variables de nivel individual, agregadas a nivel comunal ($X_{..k}$) sobre el promedio comunal de hijos; β_2 es el término correspondiente a los efectos de las variables de nivel distrital que han sido agregadas a nivel comunal ($U_{.k}$) sobre la paridez comunal; β_3 es el coeficiente utilizado para expresar los efectos de las variables propiamente comunales (V_k) sobre el promedio de paridez comunal; v_k es el término que se utiliza para expresar el error a nivel comunal, esto es, la variación de cada comuna con respecto al promedio ajustado por la ecuación.

c) *El modelo micro o análisis intradistrital*

Este análisis adopta dos decisiones iniciales: primero, todas las variables son centradas, esto es, se expresan como desviaciones de sus medias; segundo, que éste es el único nivel donde es justificable insertar alguna interacción entre predictores micro y predictores macro. La fórmula de estimación fue:

$$\log Y_{ijk} = \beta_1 X_{ijk} + e_{ijk}$$

En esta ecuación micro, $\log Y_{ijk}$ corresponde a la desviación de la paridez individual con respecto a la paridez distrital. El término β_1 es

el coeficiente de X_{ijk} , variable que corresponde a la desviación entre el valor de un predictor de nivel individual con respecto al promedio distrital de ese predictor. Este coeficiente resume entonces los efectos individuales de la variable explicativa sobre la dependiente. El término e_{ijk} concentra la variación individual con respecto al promedio de las desviaciones de paridez calculadas.

4. Procesamiento de los datos

El procesamiento de los datos² comenzó mediante la construcción de las bases de datos para el análisis interdistrital, intercomunal e intradistrital. El extenso y complicado proceso de combinación de bases de datos se vio facilitado por la flexibilidad y capacidad operativa del sistema SAS para estaciones de trabajo.

Posteriormente, los archivos fueron transformados al formato ASCII,³ para su procesamiento definitivo con *Generalized Linear Interactive Modelling* (GLIM) y MLn (*Multilevel modelling n-level of analysis*). El software nombrado fue seleccionado principalmente por su flexibilidad para la modelización estadística, especialmente para variables cuya distribución de error es distinta de la normal (Little, 1978; Pullum, 1987; Goldstein, 1995). GLIM y MLn tienen capacidades para ajustar variables dependientes con diferentes estructuras de error. Por ejemplo, en el caso de la distribución de Poisson o de la distribución binomial, la desviación (*deviance*) corresponde al valor de la razón de verosimilitud estadística (*likelihood ratio*) para el modelo ajustado comparado con el modelo saturado, que tiene un parámetro para cada variable en el modelo (Aitkin y otros, 1989). Este solo hecho sugiere que GLIM y MLn son los paquetes estadísticos más apropiados para los modelos involucrados en este tipo de investigaciones. Para la realización del presente estudio, gran parte de estos modelos fueron estimados en el laboratorio de computación del Population Research Center (PRC) de la Universidad de Texas, debido a la inexistencia de este software en la red de computación de la Universidad de Concepción.

² Para ello se utilizó el Sistema de análisis estadístico (SAS) en el ambiente de estaciones de trabajo IBM/AIX Risc 6000 de la Dirección de Planificación e Informática de la Universidad de Concepción.

³ Sistema normalizado de los Estados Unidos para el intercambio de información (American Standard Code for Information Interchange).

III. RESULTADOS

Como se dijo antes, este estudio está basado en información del Censo Nacional de Población y Vivienda de 1992 correspondiente al sector rural de la Región del Biobío. Ese año, la región comprendía 49 comunas, y el sector rural considerado en este estudio abarcaba 348 distritos rurales. Las mujeres en edad fértil (de 15 a 49 años) que vivían en la región y que forman la unidad de análisis de nivel individual (nivel 1) de este estudio, llegaban a 91 327, repartidas en 66 708 hogares ubicados en 64 926 viviendas, distribuidas a su vez en los 348 distritos rurales que constituyen uno de los niveles macro de este estudio (nivel 2) los distritos están incluidos jerárquicamente en el segundo nivel macro de este estudio, esto es, en las 49 comunas (nivel 3).

Si bien los distritos rurales y las comunas son entidades político-administrativas antes que categorías sociológicas, fueron seleccionados por diversas razones en cuanto contextos sociales pertinentes. En primer lugar, porque existe una relación jerárquica espacial entre todos los niveles incluidos en este estudio: las mujeres individuales residen en los distritos, que a su vez forman parte de las comunas. Por tanto, el requisito de anidamiento jerárquico entre las unidades de análisis está claramente establecido. Segundo, en alguna medida los distritos rurales son el equivalente sociológico rural de la noción de vecindario urbano, pues, haciendo las diferencias del caso, sus habitantes comparten el criterio de proximidad espacial relativa y, por consiguiente, de ambientes físicos y sociales. Tercero, aunque la localidad rural es una entidad censal más próxima al nivel individual, resultó excesivamente atomizada por las restricciones que imponía este estudio a la unidad de análisis de nivel individual; por ejemplo, la exclusión de las mujeres solteras o nunca casadas de entre 15 y 49 años. Cada localidad rural tuvo en promedio cuatro o cinco casos con estas características, cantidad absolutamente insuficiente para hacer mediciones robustas o, en su defecto, mínimamente estables. El distrito rural, en cambio, sin ser espacialmente tan próximo, tiene la cercanía suficiente como para asegurar la exposición a servicios comunes y proporciona además un número de casos adecuado a los propósitos del estudio. La agrupación administrativa territorial de nivel local con más claro nivel sociológico es la comuna, que ha permitido muy útiles comparaciones sociodemográficas en múltiples estudios, actuales y pasados.

La información censal original estaba disponible en tres bases de datos computacionales del Instituto Nacional de Estadísticas (INE), rela-

tivas a características de las viviendas, de los hogares y de los individuos. El procesamiento de los datos supuso la combinación de esos archivos originales para construir las bases de datos de nivel individual y contextual que requirió este estudio.

1. Datos de nivel individual

La base de datos original de nivel individual constaba de la información censal relativa a 91 327 mujeres de 14 a 49 años, distribuidas espacialmente por provincia, comuna, distrito, localidad, número de vivienda y de hogar. Estos datos de clasificación espacial se complementan con información demográfica (edad, estado civil, hijos nacidos vivos, hijos vivos, fecha de nacimiento del último hijo nacido vivo), social (parentesco, educación, religión, etnia, residencia), ocupacional (situación laboral, tipo, régimen y establecimiento de trabajo) y de incapacidad física o mental (ceguera, sordera, mudez, parálisis, deficiencia mental).

Para formar la unidad de análisis correspondiente a los patrones de fecundidad marital de las mujeres fértiles, se excluyó del estudio, como se indicó recién, a las mujeres que se declararon solteras o nunca casadas. Debido al bajo número de casos de las categorías de incapacidad física o mental, también se excluyeron estos casos, con lo cual el número definitivo de mujeres fértiles se redujo a 55 905. En la parte superior del cuadro 1 figura un resumen de los datos estadísticos referidos a las variables de nivel individual utilizadas en el estudio.

2. Datos de nivel contextual

A partir de los registros de viviendas y hogares del INE, se construyó una base de datos contextual, combinada, con las proporciones de las características de la vivienda (tipo, tenencia, alumbrado, origen y ubicación del agua, baños, ducha) y de los hogares (tipo de combustible, disponibilidad de radio, televisor, vídeo, equipo de música, lavadora, refrigerador, horno microondas, teléfono, bicicleta, automóvil) en los 348 distritos rurales existentes en las 49 comunas de la Región del Biobío.

Es importante observar que la base de datos de nivel macro o contextual permite el análisis entre distritos, porque agrega las variables de nivel hogar y de nivel vivienda existentes en los 348 distritos rurales. También permite el análisis entre comunas, porque concentra jerárquicamente la información censal original de los hogares y de las viviendas de las mujeres de las 49 comunas estudiadas. En la segunda parte del

Cuadro 1
**ESTADÍSTICAS DESCRIPTIVAS DE LAS VARIABLES DE NIVEL
INDIVIDUAL Y CONTEXTUAL DE LAS MUJERES CASADAS
DEL SECTOR RURAL DE LA REGIÓN DEL BIOBÍO, 1992**

Variable	Definición	Media aritmética	Desviación estándar
Nivel individual			
Edad	Edad en años calendario	33.29	8.66
Número de hijos	Paridez total de la mujer	3.14	2.17
Antecedentes migratorios	Migración previa (0=no, 1=sí)	0.33	0.47
Educación	Años de educación	5.95	3.32
Estado civil	Casada (0=no, 1=sí)	0.96	0.20
Ocupación	Sólo trabaja en casa (0=no, 1=sí)	0.92	0.27
Religión	Es católica (0=no, sí=1)	0.67	0.47
	Es pentecostal (0=no, 1=sí)	0.28	0.45
Nivel contextual			
Tipo de vivienda	Porcentaje de casas	0.86	0.12
	Porcentaje de mejoras	0.14	0.12
Tenencia	Porcentaje de propietarios	0.56	0.19
	Porcentaje de casas cedidas	0.40	0.19
Alumbrado	Porcentaje con electricidad	0.37	0.26
Origen del agua	Porcentaje red pública	0.13	0.20
	Porcentaje con pozo	0.50	0.31
	Porcentaje otros tipos	0.37	0.32
Abastecimiento interno de agua	Porcentaje con agua en la vivienda	0.19	0.16
Tipo de servicio higiénico	Porcentaje con alcantarillado	0.11	0.13
	Porcentaje con pozo negro	0.89	0.13
Tipo de combustible para cocinar	Uso de leña (0=no, 1=sí)	0.86	0.15
Existencia de ducha	(0=no, 1=sí)	0.14	0.13
Radio	(0=no, 1=sí)	0.79	0.08
Televisor en blanco y negro	(0=no, 1=sí)	0.42	0.18
Televisor en color	(0=no, 1=sí)	0.11	0.13
Vídeo	(0=no, 1=sí)	0.02	0.05
Equipo de música	(0=no, 1=sí)	0.06	0.10
Lavadora	(0=no, 1=sí)	0.07	0.10
Refrigerador	(0=no, 1=sí)	0.13	0.14

Fuente: Elaborado por el autor, sobre la base del Censo Nacional de Población y Vivienda de 1992.

cuadro 1 se presentan las estadísticas descriptivas de la información de nivel contextual del estudio.

En las bases de datos de nivel contextual (tanto para el análisis entre distritos como para el análisis entre comunas) se agregó la variable dependiente de nivel individual del estudio, esto es, el número total de hijos vivos tenidos por las mujeres en edad fértil de la región. Obviamente, esa variable fue procesada y agregada como proporción tanto a nivel distrital como a nivel comunal.

3. Selección de predictores significativos

El proceso de modelización de niveles múltiples comienza con la selección de variables predictoras y el desarrollo de modelos, mediante la aplicación de las técnicas utilizadas en el análisis estadístico corriente de un solo nivel. La idea es realizar el proceso de selección de todos los predictores que son significativos en cuanto a reducir la variación de la variable dependiente en cada uno de los niveles considerados en el análisis. El procedimiento técnico habitual consiste en examinar las relaciones de orden cero entre la variable dependiente y cada uno de los predictores en los diferentes niveles considerados en el estudio.

En este caso se hizo un examen aún más exigente. Como en los últimos años las pruebas de significación estadística y los valores p que les están asociados en la inferencia estadística han presentado cada vez más problemas prácticos y resultados anómalos cuando se trata de estudios con muestras grandes (existe evidencia de que en tal caso los valores p tienden a indicar rechazo de la hipótesis nula, aun cuando el modelo nulo pareciera teóricamente más razonable que el alternativo), se ha utilizado aquí la aproximación conocida como *Bayesian Information Criteria* (BIC), del enfoque estadístico bayesiano, para la seleccionar los modelos unidimensionales. Este procedimiento, de creciente uso en sociología desde 1986, evita los problemas de influencia de los valores p por el tamaño de la muestra y permite una selección de modelos más racional en la investigación social (Raftery, 1994; Kass y Raftery, 1995).

La selección de modelos bayesiana ha sido introducida en la investigación social por Adrian Raftery, por lo que remitimos a sus escritos para un desarrollo teórico y empírico de la verificación de hipótesis en este importante paradigma estadístico (Raftery 1986, 1988, 1993, 1994). Para nuestros efectos sólo es importante establecer que la aproximación BIC es un procedimiento cuantitativo que permite estimar

el factor de Bayes mediante una simple ecuación, que se ajusta a diferentes clases específicas de variables dependientes. Para el caso general la fórmula es:

$$BIC_k = L_k^2 - df_k \log n,$$

donde L_k^2 es la desviación para el modelo M_k y df_k es el número correspondiente de los grados de libertad del modelo. BIC_s es el valor BIC para el modelo saturado y es igual a 0. El modelo saturado es preferible a M_k si $BIC_k > 0$, caso en el cual M_k es considerado un modelo que no se ajusta bien a los datos. Cuando $BIC_k < 0$, M_k es preferible al modelo saturado, y mientras más pequeño sea BIC_k (esto es, mientras más negativo), mejor es el ajuste del modelo M_k .

Cuando el modelo de comparación inicial (*baseline*) es el modelo nulo, M_0 , aquel que no tiene ninguna variable independiente, entonces BIC_k es reemplazado por BIC'_k . Para los modelos de regresión lineal la formulación más conveniente de BIC' , es la siguiente:

$$BIC'_k = n \log(1-R_k^2) + p_k \log n,$$

donde R_k^2 es el valor de R^2 (coeficiente de determinación) para el modelo M_k y p_k es el número de variables independientes (sin incluir el intercepto). De la misma manera que en el caso general, si BIC'_k es positivo, debe preferirse el modelo nulo, lo que indica que el modelo alternativo está sobreparametrizado. Pero si BIC'_k es negativo, entonces el modelo alternativo es preferible al modelo nulo, y mientras más pequeño sea el valor del BIC' (esto es, mientras más negativo), más preferible es el modelo.

La evidencia de la significancia de una variable independiente adicional puede ser medida por:

$$BIC'_{k+1} - BIC'_k = n \log \left\{ \frac{(1 - R_{k+1}^2)}{(1 - R_k^2)} \right\} + \log n$$

Aquí M_k está incluido jerárquicamente en M_{k+1} , que contiene una variable adicional. Para que haya evidencia en favor de la nueva variable, la ecuación anterior debería entregar un resultado negativo. En este estudio, estos criterios de evidencia son utilizados en el análisis del cuadro 5.

Como la modelización que se utilizó aquí para seleccionar variables predictoras significativas del número de hijos contrasta la variación total del modelo nulo con la de modelos que agregan sucesi-

vamente variables predictoras independientes, se debe utilizar BIC' como la aproximación bayesiana más apropiada para la selección de los modelos pertinentes.

4. Predictores de nivel individual

Una excelente ilustración de la eficacia del procedimiento BIC' en la selección de modelos pertinentes está dada en la selección de los predictores de nivel individual. Cuando los predictores censales de nivel individual fueron seleccionados mediante los procedimientos usuales del análisis de datos, que consisten en utilizar los valores p de la prueba F para comparar coeficientes de determinación en modelos de regresión con diversas variables independientes, se encontró que todas las variables de nivel individual, excepto la etnia de la mujer, tenían relaciones bivariadas altamente significativas con la variable respuesta. Esto puede ser atribuido fundamentalmente a la influencia del gran tamaño de la población censal utilizada.

Cuando se aplicó el procedimiento BIC' para detectar predictores significativos, se encontró que a nivel individual la educación y la edad de la mujer fértil tenían efectos sobre el número total de hijos considerablemente superiores al detectado para las otras variables. Como puede observarse en el cuadro 2, de los siete predictores utilizados, éstas son las dos variables que alcanzan un valor de BIC' negativo más marcadamente diferente, lo que revela que ambas explican la mayoría de la variación de la variable dependiente.

Es fácil ver que, a nivel individual la edad de la mujer explica un 30.62% de la variación del número total de hijos. Sin embargo, en este estudio la edad de la mujer es usada no como un predictor sino como una variable control, para captar, al menos parcialmente, la exposición de la mujer al riesgo de concebir, en ausencia de información censal relativa a la duración del matrimonio. La educación de la mujer, medida como una variable categórica, es el predictor más relevante del modelo, porque da cuenta de una parte sustancial de la variación de la variable respuesta. Los cinco estratos de educación utilizados explican el 15.48% del total de la variación del número total de hijos tenidos por la mujer rural. En el cuadro 2 se ha utilizado como categoría educacional de referencia el estrato de las mujeres rurales sin educación formal (edu1). El segundo estrato de educación (edu2) reúne a las mujeres que declaran entre uno y tres años de escolaridad formal; el tercero (edu3), a las que declaran entre cuatro y seis años; el cuarto (edu4), a las que declaran

Cuadro 2
**SELECCIÓN DE VARIABLES DE NIVEL INDIVIDUAL
 CON EFECTOS DE ORDEN CERO SIGNIFICATIVOS
 SOBRE EL NÚMERO DE HIJOS**

Variables	R ²	Nº de categorías	N	BIC'
Edad	.3062	6	55905	-20 372
Educación	.1548	4	55905	-9 359
Antecedentes migratorios	.0001	1	55905	5.3
Estado civil	.0102	1	55905	-281
Ocupación	.0108	3	55905	-180
Situación ocupacional	.0038	2	55905	-174
Religión	.0034	2	55905	-83
Etnia	.0001	2	55905	16.3
Actividad	.0032	1	55905	-168

Fuente: Elaborado por el autor, sobre la base del Censo Nacional de Población y Vivienda de 1992.

entre siete y nueve años, y el quinto (edu5) a las que tienen más de 10 años de educación formal.

Sólo las variables descriptoras de los antecedentes migratorios y de la etnia de la mujer no resultaron asociadas en términos bivariados al número de hijos y fueron, por tanto, eliminadas del análisis posterior de nivel individual. El estado civil y la ocupación explican, cada uno, algo más de 1% de la variación de la variable respuesta, mientras que la religión, el tipo de ocupación y de actividad laboral demuestran relaciones aún más residuales. En resumen, dentro del conjunto de los predictores de nivel individual, sólo la edad, la educación, la ocupación, la religión y el estado civil tienen efectos estadísticos significativos sobre el número total de hijos. Ninguna otra variable, pese a obtener valores significativos en sus pruebas *F* en modelos de regresión bivariada, alcanzó valores negativos en la aproximación BIC', por lo que a nivel individual fueron desechadas de las futuras modelizaciones. Del conjunto de variables significativas expuestas en el cuadro 2 fueron eliminadas, además, la situación ocupacional y la actividad, debido a que surgen problemas de colinealidad con el tipo de ocupación. Por tanto, en la determinación del modelo óptimo de nivel individual sólo se consideraron las variables edad, educación, estado civil, tipo de ocupación y religión.

Para determinar el modelo óptimo de nivel individual se procedió a comparar los 32 modelos posibles que comprenden desde una a las cinco variables simultáneamente. El resumen de ese ejercicio puede ser examinado en el cuadro 3.

Cuadro 3
SELECCIÓN DE MODELO ÓPTIMO A NIVEL 1 (INDIVIDUAL)

Modelos	R ²	Nº de categorías	BIC'
A	.3062	6	-20 372
E	.1548	4	-9 358
V	.0052	1	-280
O	.0038	3	-180
R	.0019	2	-84
A E	.3519	10	-24 137
A E R	.3558	12	-24 453
A E R O	.3579	15	-24 603
A E R O V	.3580	16	-24 600

Fuente: Elaborado por el autor, sobre la base del Censo Nacional de Población y Vivienda de 1992.

Nota: El modelo óptimo es el que aparece en negritas. Las variables independientes son:

- A = edad de la mujer (7 categorías)
- E = educación (5 categorías)
- V = estado civil (0 = sin pareja; 1 = con pareja)
- O = ocupación (4 categorías)
- R = religión (3 categorías)

El modelo óptimo a nivel 1 comprende sólo cuatro variables: la edad, la educación, la ocupación y la religión de las mujeres. Este modelo tiene un valor BIC' de -24 603, el más alto de los 32 modelos considerados para las cinco variables seleccionadas a nivel individual. Esto significa que en este nivel y para las variables consideradas en la información censal, la paridez de las mujeres rurales es esencialmente atribuible a esas cuatro variables (edad, educación, ocupación y religión). El modelo de cuatro variables seleccionado explica el 35.8% de la variación total de la variable dependiente. Estas cuatro variables fueron las representantes del nivel 1 que se utilizaron en la modelización de niveles múltiples posterior.

5. Predictores contextuales

a) *Variables a nivel de distrito*

Las variables contextuales de este estudio estuvieron mayoritariamente constituidas por las características de los hogares y de las viviendas de las mujeres fértiles encuestadas en el censo de 1992. En lo que se refiere a los hogares, el censo proporciona esencialmente información de naturaleza socioeconómica, mediante indicadores como el tipo de combustible más empleado para cocinar y la existencia de una

serie de artículos eléctricos, como radio, televisor, refrigerador, o lavadora. En cuanto a las características de la vivienda, el censo recoge datos acerca de la existencia y calidad de los servicios básicos de electricidad, agua potable, alcantarillado y eliminación de excretas. En este estudio, un paso previo al análisis de niveles múltiples consistió en la construcción de una base de datos combinada de los hogares, que incluye información sobre los servicios básicos de cada vivienda, más la disponibilidad de enseres de cada hogar. Ese archivo se agregó a nivel de los 348 distritos rurales mediante el cálculo de proporciones de cada variable en el distrito considerado. Para tales efectos se convirtió cada variable en un conjunto de indicadores (de rango 0-1), y se agregó al archivo la proporción de casos de la categoría. Esto permite la agregación de cada variable a nivel de distrito (nivel 2 de este estudio) y la inclusión de estas variables categóricas en ecuaciones de regresión. Se agregaron además a esta base de datos de nivel 2 los promedios distritales de edad, educación, antecedentes migratorios, estado civil, religión, tipo y nivel ocupacional, todas características composicionales provenientes del archivo de nivel individual. Sólo se excluyó de esta operación, destinada a medir posibles efectos contextuales, la etnia de la mujer, ya que esta variable no presenta suficiente variabilidad en la región. En el cuadro 4 se exhiben las relaciones bivariadas de cada predictor con la variable dependiente a nivel de distrito.

Como se advierte en el cuadro 4, nueve variables tienen algún efecto sobre la variable respuesta, lo cual se expresa en porcentajes relativamente importantes de explicación de su variación y en valores BIC' negativos. Obsérvese que la proporción de educación en el distrito da cuenta del 37.7% de la variación de la paridez distrital total. Su valor BIC', igual a -160.5, muestra que es con mucho una de las variables predictoras con mayores efectos independientes a nivel de distrito. Debe también prestarse atención aquí al comportamiento de la variable (promedio distrital de edades), pues mientras esta variable tiene efectos considerables a nivel individual, a nivel agregado distrital su efecto se reduce a 43.6% de la variación de la paridez marital y alcanza además el mismo valor BIC' que la anterior, lo que significa que en este nivel contextual también es la variable más importante en la magnitud del efecto de orden cero sobre la respuesta. Otros efectos importantes son el 5.3% debido al tipo de trabajo en el distrito, el 6.1% que muestra la proporción de hogares donde hay bicicleta, el 4.6% debido a la proporción de hogares con vídeo, o el 2% correspondiente al tipo de alcantarillado existente.

Cuadro 4
**SELECCIÓN DE VARIABLES DE NIVEL DISTRITAL CON EFECTOS
SIGNIFICATIVOS DE ORDEN CERO SOBRE EL PROMEDIO
DE PARIDEZ EN EL DISTRITO**

Variable	R ²	Nº de categorías	N	BIC'
Edad	.4364	1	348	-164.4
Educación	.3767	1	348	-160.5
Antecedentes migratorios	.1350	1	348	-49.5
Estado civil	.0111	1	348	2.0
Trabajo	.0530	1	348	-1.4
Ocupación	.0510	1	348	-17.2
Religión	.0131	1	348	7.1
Tipo de vivienda	.0016	1	348	5.3
Tenencia	.0000	1	348	5.8
Alumbrado	.0020	1	348	5.2
Abastecimiento de agua	.0065	1	348	3.6
Agua interior	.0042	1	348	4.4
Tipo de combustible para cocinar	.0131	1	348	1.3
Existencia de alcantarillado	.0192	1	348	-0.9
Ducha	.0013	1	348	5.4
Radio	.0167	1	348	0
Televisor en blanco y negro	.0062	1	348	4.1
Televisor en color	.0050	1	348	4.1
Vídeo	.0456	1	348	-10.4
Equipo de música	.0382	1	348	-7.7
Lavadora	.0059	1	348	3.8
Refrigerador	.0029	1	348	4.8
Bicicleta	.0609	1	348	-16.0

Fuente: Elaborado por el autor, sobre la base del Censo Nacional de Población y Vivienda de 1992.

Nota: Los predictores significativos son los que aparecen en negritas.

A nivel 2 ó distrital, de las nueve variables con efectos significativos de orden cero sobre la respuesta, cinco son promedios distritales de variables de nivel individual (edad, educación, antecedentes migratorios, tipo y nivel de ocupación), una corresponde a características distritales de las viviendas (tipo de alcantarillado), y tres caracterizan al hogar de la mujer en el distrito (disponibilidad de vídeo, de equipo de música y de bicicleta). Al buscar un modelo óptimo entre las cinco características agregadas desde el nivel individual, se concluyó que sólo la edad y la educación captan la totalidad de los efectos de este origen, por lo que en este nivel 2 puede prescindirse de los antecedentes migratorios o de los rasgos laborales. De las tres variables del hogar, debe eliminarse la disponibilidad de equipo de música, rasgo que es captado adecuadamente por la existencia de vídeo y de bicicleta. A nivel de vivienda, se mantuvo en el modelo la variable referida al tipo de alcantarillado.

Cuadro 5
SELECCIÓN DEL MODELO ÓPTIMO A NIVEL 2 (DISTRITAL)

Modelos	R ²	Nº de categorías	BIC'
A	.4364	6	-164.4
E	.3767	4	-160.5
C	.0192	1	-0.9
V	.0456	1	-10.4
B	.0609	1	-16.0
A E	.6356	10	-292.9
A E C	.6784	11	-330.4
A E C V	.6889	12	-336.1
A E C V B	.6976	13	-340.1

Fuente: Elaborado por el autor, sobre la base del Censo Nacional de Población y Vivienda de 1992.

Nota: El modelo óptimo es el que aparece en negritas. Las variedades independientes son:

- A = edad de la mujer (7 categorías)
- E = educación de la mujer (5 categorías)
- C = proporción distrital de viviendas con alcantarillado
- V = proporción distrital de hogares con vídeo
- B = proporción distrital de hogares con bicicleta.

El paso siguiente es establecer un modelo óptimo final a nivel 2. Esto significa examinar la posibilidad de que a partir de las cinco variables predictoras encontradas se pueda construir un modelo más acorde con el principio de parsimonia, esto es, que junto con explicar también significativamente la paridez distrital, lo haga de un modo más simple y eficiente.

Como puede observarse en el cuadro 5, donde se expresan los modelos propuestos y sus resultados, el modelo óptimo al nivel distrital incorpora sólo los efectos de tres variables: el promedio de edad, de educación y de hogares con alcantarillado. Este modelo tiene un valor BIC' de -330.4, y explica un 67.8% de la variación de la paridez marital distrital. Sin embargo, como se ve en el cuadro 5, entre los 32 modelos ajustados a nivel distrital, hay dos (uno de cinco variables y uno de cuatro) que tienen un valor BIC' inferior al seleccionado como óptimo. Por lo tanto, conviene aclarar por qué se ha preferido en este caso un modelo con mayor valor BIC'. La explicación es algo compleja. Raftery (1994) sostiene que al contrastar dos modelos jerárquicos para observar la influencia de una variable específica, deben considerarse sus diferencias en valores BIC'. Según Raftery, cuando las diferencias entre ambos modelos son de hasta cinco puntos, sólo permiten una muy leve indicación del probable efecto de la variable analizada. Cuando hay entre

5 y 10 puntos de diferencia, la evidencia es un poco más sostenible, y se llega a una evidencia concluyente de los efectos específicos de la variable analizada cuando la diferencia es de más de 10 puntos.

Estas diferencias de valores BIC' otorgan distintos grados de evidencia a la inclusión de una variable adicional. Aunque estos puntajes BIC' se utilizan para eliminar las incongruencias de los valores probabilísticos que presentan las pruebas de significación en investigaciones que utilizan muestras grandes, se puede demostrar que tales grados corresponden cercanamente a valores de probabilidad muy pequeños, lo que hace altamente probable que la decisión acerca del parámetro sea correcta (Raftery, 1994, pp. 22 a 24). En este estudio se ha utilizado el criterio más exigente como evidencia del aporte de una variable adicional. Por lo tanto, sólo se aceptaron variables adicionales que implican una diferencia superior a 10 puntos entre los valores BIC' de los modelos jerárquicos comparados.

Al interpretar el cuadro 5 en este contexto, surge como modelo óptimo aquel que incluye tres variables con un valor BIC' de -330.4 y una reducción del 67.8% de la variación de la variable dependiente. Aunque el mejor modelo desde el punto de vista del valor más pequeño de BIC' es aquel que reúne las cinco variables (BIC' = -340.1), su diferencia con el mejor modelo de cuatro variables es de sólo cuatro puntos, lo que indica que la evidencia de un efecto significativo para la quinta variable no es sustancial. En tales circunstancias y de acuerdo con el principio de parsimonia, es preferible el modelo de cuatro variables, el cual, con un valor BIC' de -336.1, explica un 68.9% de la variación de la variable dependiente. Sin embargo, al comparar el modelo de cuatro variables con aquel mejor de sólo tres variables, nuevamente se obtiene una diferencia en valores BIC' ligeramente superior a cinco puntos. La lógica de la modelización indica, por las mismas razones anteriores, que es preferible el modelo de sólo tres variables, esto es, aquel que incorpora la edad, la educación y el promedio distrital de viviendas con alcantarillado. Este modelo tiene un valor BIC' de -330.4 y explica un 67.8% de la variable respuesta. Al comparar este modelo con el mejor de dos variables se advierte que la diferencia entre ambos es de más de 30 puntos, diferencia atribuible por entero al efecto de la tercera variable (C en el modelo AEC con respecto al modelo AE). Esta diferencia es absolutamente evidente y no puede ser desechada, por lo que la decisión debe ser seleccionar el modelo AEC como óptimo.

Al seleccionar el modelo óptimo desde el mejor modelo de una variable (A), se llega a idénticos resultados. En el modelo de dos variables

(AE) el valor de BIC' disminuye a -292.8, una diferencia de más de 120 puntos con respecto al mejor modelo univariado. Al comparar con el mejor modelo de tres variables (AEC), el valor de BIC' disminuye a -330.4, esto es, más de 30 puntos con respecto al modelo bivariado. Ciertamente, el modelo de tres variables es el mejor. Al compararlo con el de cuatro términos, la disminución es de sólo cinco puntos. Esto significa que no hay evidencias sólidas de que la cuarta variable contribuya significativamente al modelo. De nuevo la conclusión es que AEC, el modelo formado por la edad, la educación y la proporción de hogares con alcantarillado, es el óptimo a nivel del distrito rural.

En conclusión, en el nivel 2 de este estudio (esto es, en los 348 distritos incluidos en las 49 comunas de la Región del Biobío), las variables que influyen significativamente sobre la paridez son los promedios de educación, de edad y de viviendas con alcantarillado.

b) *Variables de nivel comunal*

El segundo nivel contextual utilizado en este estudio corresponde a las comunas en que están jerárquicamente situados los distritos rurales. Las comunas constituyen, por lo tanto, el nivel 3 de esta investigación. La agregación en este nivel de las variables censales de hogares y viviendas permitió construir un archivo con 49 registros, que reúne los promedios comunales referentes a la disponibilidad de servicios básicos en la vivienda y a las dimensiones socioeconómicas representadas por la existencia de aparatos electrónicos en el hogar. La variable paridez marital comunal fue agregada al archivo como el promedio del total de hijos tenidos por las mujeres del sector rural de cada comuna, a partir de la variable de nivel individual constituida por el número total de hijos que declaró cada una de ellas en el censo analizado.

Las regresiones bivariadas entre cada predictor comunal y la respuesta en el mismo nivel permitieron constatar que ocho variables de la base de datos tienen efectos significativos sobre la variable dependiente. En el cuadro 6 se muestran los resultados del análisis de datos bivariado a nivel comunal:

Un primer hecho destacable en las relaciones de orden cero a nivel comunal consiste en que el promedio de edad no tiene aquí efectos significativos sobre la variable respuesta. Curiosamente, la edad de las mujeres no es una variable relacionada con la paridez en el nivel contextual de la comuna. No obstante, tal como en el nivel 2, la educación está también estrechamente asociada aquí a la paridez, como lo indica su valor BIC', de -27.1. El promedio comunal de educación explica

Cuadro 6
**SELECCIÓN DE VARIABLES DE NIVEL COMUNAL CON EFECTOS
SIGNIFICATIVOS DE ORDEN CERO SOBRE EL PROMEDIO
DE PARIDEZ EN LA COMUNA**

Variable	R ²	Nº de categorías	N	BIC'
Edad	.2368	6	49	10.1
Educación	.5810	4	49	-27.7
Antecedentes migratorios	.1068	1	49	-1.6
Estado civil	.0053	1	49	3.6
Trabajo	.0266	3	49	10.4
Ocupación	.0887	2	49	3.2
Religión	.0395	2	49	5.8
Tipo de vivienda	.2388	1	49	-9.5
Tenencia de la vivienda	.0437	1	49	1.7
Alumbrado	.2647	1	49	-11.2
Abastecimiento de agua	.1544	1	49	-4.3
Ubicación del agua	.2462	1	49	-9.9
Tipo de combustible para cocinar	.0367	1	49	-14.1
Existencia de alcantarillado	.3115	1	49	-14.4
Ducha	.3826	1	49	-19.7
Radio	.0710	1	49	0.3
Televisor en blanco y negro	.2241	1	49	-8.5
Televisor en color	.3782	1	49	-19.4
Vídeo	.3089	1	49	-14.2
Equipo de música	.3647	1	49	-18.3
Lavadora	.3759	1	49	-17.8
Refrigerador	.3899	1	49	-20.3
Bicicleta	.0154	1	49	3.1

Fuente: Elaborado por el autor, sobre la base del Censo Nacional de Población y Vivienda de 1992.
Nota: Las variables significativas están destacadas en negritas.

58.1% de la variación del promedio comunal de paridez. El examen del cuadro 6 revela efectos provenientes de 15 variables. Sólo dos variables agregadas a nivel comunal a partir de las características individuales resultaron significativas: la educación y el promedio de personas con antecedentes migratorios. Siete de las ocho variables de nivel vivienda resultaron significativas en el nivel 3, en el cual sólo la variable tenencia de la vivienda no guarda relación con la paridez comunal. Otras seis características del hogar resultaron predictores independientes significativos de la paridez en el análisis inicial al nivel de comuna. Los promedios de televisores en la comuna, de vídeos, de equipos de

música, de lavadoras y de refrigeradores explican, cada uno, entre 30.9% y 38.9% de la variación del resultado, con valores BIC' altamente significativos, ubicados entre -8.5 y -20.3. Esto significa que a nivel comunal las relaciones bivariadas indican la existencia de varios efectos importantes sobre la paridez, mientras que en el nivel 2, o distrital, sólo se detectaron efectos de la edad, la educación y de la existencia de alcantarillado.

Al modelar los efectos combinados y simultáneos de los predictores comunales que tienen relaciones significativas con la variable dependiente, y en la perspectiva de encontrar un modelo óptimo que explique los efectos sobre la paridez comunal detectados en el trabajo, en un primer paso se procedió a determinar la colinealidad de los predictores importantes a nivel de hogar. La matriz de intercorrelaciones de estos predictores reveló la existencia de una elevada correlación entre ellos, por lo que se estimó pertinente eliminar las variables representadas por la proporción de vídeos, de equipos de música, de televisores y lavadoras, y mantener en el modelo sólo la variable restante, esto es, la proporción de refrigeradores, dado que ésta capta por sí sola casi todos los efectos detectados para las variables eliminadas. En el caso de las variables descriptivas de servicios básicos de la vivienda, se determinó, por las mismas razones, que las variables que mejor captaban los efectos de las otras eran la proporción de tipos de vivienda, la proporción de viviendas con electricidad y la proporción de casas con ducha interna, por lo que estas variables se dejaron en el modelo, al tiempo que fueron excluidos de él, por multicolineales, el promedio comunal de casas con alcantarillado y el promedio comunal de casas que usan gas como combustible principal. En resumen, las variables que captan los efectos del contexto comunal sobre la paridez en el nivel 3 resultaron ser: los promedios comunales de educación, de movilidad, de casas con luz, con ducha y con refrigerador.

El siguiente paso de la modelización implicó estimar ecuaciones de regresión entre los predictores comunales que mejor se ajustaban a la explicación de la paridez agregada a este nivel. El propósito era seleccionar el modelo óptimo en el nivel 3, que explica con más parsimonia la variable respuesta del estudio. En el cuadro 7 se muestra el proceso de selección, basado en aproximaciones al factor de Bayes mediante el uso de BIC'.

El modelo óptimo en el nivel 3 es aquel que comprende sólo los efectos de la variable constituida por el promedio de educación a nivel comunal, cuyo valor BIC' (-27) es el más bajo de todos los modelos

Cuadro 7
SELECCIÓN DE MODELO ÓPTIMO FINAL
A NIVEL 3 (COMUNAL)

Modelos	R ²	P	BIC'
E	.5810	4	-27
M	.1068	1	-2
C	.2388	1	-10
L	.2647	1	-11
D	.3826	1	-20
R	.3899	1	-20
E R	.6141	5	-27
E R L	.6212	6	-24
E R L M	.6246	7	-21
E R L M C	.6284	8	-17
E R L M C D	.6285	9	-14

Fuente: Elaborado por el autor, sobre la base del Censo Nacional de Población y Vivienda de 1992.

Nota: El modelo óptimo es el que aparece en negritas. Las variables independientes son las siguientes:

- E = promedios comunales de educación (5 categorías)
- M = promedio comunal de personas migrantes (1 = sí; 0 = no)
- C = promedios comunales de tipos de vivienda (1 = casa; 0 = otros)
- L = promedio comunal de viviendas con electricidad
- D = promedio comunal de viviendas con ducha
- R = promedio comunal de viviendas con refrigerador

comparados, incluso, por supuesto, de aquel que reúne simultáneamente todas las variables con efectos independientes de orden cero. Este último tiene un valor BIC' igual a -14, lo que está señalando un mal ajuste en comparación con el modelo que abarca sólo la educación. De los modelos con dos variables, el mejor es aquel que reúne simultáneamente la educación y la existencia de refrigeradores. Su valor es de -27 en la aproximación BIC', el mismo que el del modelo que sólo comprende la educación, a pesar de la introducción de la variable adicional representada por la presencia de refrigeradores. Esto significa que el modelo que incorpora sólo la educación es preferible al de dos variables, en virtud del principio de parsimonia: a igualdad de valores, debe preferirse aquel modelo cuya ecuación tenga menos términos y que, por tanto, sea más simple y de más fácil interpretación.

En resumen, en el nivel 3 sólo es significativa la variable proporción comunal de los distintos grupos educacionales. Se agregará también aquí el promedio de edad, a pesar de su no significación estadística en este nivel, con el propósito de mantener su función como variable control, en ausencia de otras variables que midan la exposición de las mujeres al riesgo de concebir.

IV. REGRESIÓN DE POISSON DE NIVELES MÚLTIPLES DE LA FECUNDIDAD MARITAL

El análisis de niveles múltiples comienza por determinar qué proporción de la variación de la variable dependiente de nivel individual puede ser explicada por cada uno de los niveles agregados que la investigación establece. En este caso se trata de establecer qué proporción de la variación de la variable número total de hijos obedece a la residencia de las madres en distritos (nivel 2) y comunas (nivel 3) específicos. Ciertamente, si toda la variación fuese explicable sólo a nivel individual, no existiría justificación para realizar un análisis de niveles múltiples. En el contexto de los modelos lineales generalizados, la medida utilizada para establecer la dispersión alrededor de valores esperados se denomina desviación. En general, la desviación es -2 veces el logaritmo natural de la función de máxima verosimilitud (McCullagh y Nelder, 1983). Cuando los valores esperados corresponden a la media aritmética global, sin ningún predictor en el modelo, la desviación se llama desviación inicial. Para variables dependientes continuas, la desviación inicial es igual a la suma de las desviaciones cuadráticas con respecto a la media general. En el caso de una distribución de Poisson, como la variable dependiente de este estudio, la desviación es igual a:

$$S = \sum w_i \left\{ y_i \log \frac{y_i}{\mu_i} (y_i - \mu_i) \right\}$$

En esta fórmula, S es la desviación de la distribución de Poisson; w_i es la frecuencia o número de veces con que se repite cada valor de la variable empírica y_i con distribución de Poisson, y μ_i es la media estimada de la distribución. La fórmula de la desviación S , en el caso de una variable con distribución de Poisson, equivale a la desviación de errores cuadráticos en torno a la media de las variables con distribución normal/gaussiana (Gilchrist y Green, 1993).

En el cuadro 8 se muestran las desviaciones para los modelos de niveles múltiples que se han utilizado en esta investigación. En la columna final aparece la desviación total para la variable dependiente, y en las tres primeras columnas se muestra cómo la desviación total se divide entre aquella que es explicable a nivel individual (desviación intradistrital) y aquella que es explicable a nivel macro o contextual (desviación entre distritos y desviación entre comunas).

Cuadro 8
**DESVIACIÓN EN NÚMERO TOTAL DE HIJOS DE LAS MUJERES
 FÉRTILES DE LA REGIÓN DEL BIOBÍO: DIVISIÓN
 INTRADISTRITAL, INTERDISTRITAL
 E INTERCOMUNAL**

Modelos	Desviación intradistrital (nivel 1)		Desviación intradistrital (nivel 2)		Desviación intradistrital (nivel 3)		Desviación total
Variación total	81 178		821.84		661.00		82 661
	(98.2%)		(0.8%)		(1.0%)		(100%)
Grados de libertad (gl)	55 904		347		48		55 904
Desviación univariada debida a:							
1) Edad	26 613	6 gl	104.90	6 gl	155.24	6 gl	26 873
2) Educación	12 161	4 gl	291.00	4 gl	328.42	4 gl	12 780
3) Alcantarillado			76.80	1 gl			
4) Ocupación	322.6	3 gl					
5) Religión	151.7	2 gl					
Desviación total del modelo óptimo final:	30 301.04	15 gl	344.9	11 gl	382.8	10 gl	31 028
R² debidos a modelos univariados de:							
Edad	32.8%		12.8%		23.5%		32.5%
Educación	15.0%		35.4%		50.0%		15.5%
Alcantarillado			9.3%				
Ocupación	0.4%						
Religión	0.2%						
R² del modelo óptimo final:	37.3%		41.9%		57.9%		37.5%

Fuente: Elaborado por el autor, sobre la base del Censo Nacional de Población y Vivienda de 1992.

En la primera hilera del cuadro 8 se ofrecía que del total de variación de la variable dependiente, un 98% es explicable a nivel individual. Un 2% debe ser atribuido a la existencia de factores contextuales de nivel distrital y comunal. Aunque la proporción de efectos contextuales parece bastante pequeña, es estadísticamente significativa en ambos niveles macro. El 2% de variación de la paridez marital que es explicable en términos contextuales equivale a 821.84 puntos, que corresponden a la variación de la paridez entre distritos, y a 661 puntos, que corresponden

a la variación de la variable dependiente entre comunas. Las características de nivel 2 (348 distritos rurales) explican 1% de la variación total de la paridez marital, mientras que los segmentos de nivel 3 (49 comunas) explican el otro 0.8% de la variación contextual de la paridez marital total. El resto de la variación de la variable dependiente es atribuible sólo a características de nivel individual (nivel 1).

No es sorprendente que la magnitud de los efectos contextuales detectados en este estudio sea bastante reducida. En la literatura internacional hay muchos ejemplos de estudios empíricos, de distinto nivel, en que los efectos contextuales no suelen pasar del 10% de la variación total. Debe considerarse, además, que las agrupaciones contextuales utilizadas en el presente estudio, distritos rurales y comunas, son de carácter principalmente administrativo territorial, más que sociológico o cultural, por lo que podrían no ser las más apropiadas para captar las diferencias grupales de fecundidad.

Sin embargo, el 2% de efectos contextuales detectado es sumamente significativo desde el punto de vista estadístico, por lo que cabe concluir que corresponde a diferencias realmente existentes y merecedoras de un examen detenido. Como se puede observar en el cuadro 8, en que se sintetizan las modelizaciones de niveles múltiples, sólo las variables edad y educación resultaron significativas en los distintos niveles del estudio. Sin embargo, como se ha dicho antes, la variable edad se utiliza aquí como una variable control de la exposición de la mujer al riesgo de concebir, en ausencia de datos censales acerca de la duración marital o de la unión conyugal pertinente. Por lo tanto, el único predictor con efectos significativos en todos los niveles es la educación de la mujer fértil. Este hecho es destacable, porque estamos en una situación en que la misma variable presenta simultáneamente efectos significativos en el nivel individual y en los dos niveles contextuales modelados. Esto implica que la variable tuvo efectos significativos simultáneos e independientes a nivel individual, distrital y comunal, como se analizará en las páginas siguientes.

1. Variación de la fecundidad intradistrital o de nivel individual (nivel 1)

Éste es el nivel de investigación habitual, en que las características de los individuos son utilizadas para explicar respuestas individuales. En este caso, la respuesta individual es el número total de hijos declarado por las mujeres fértiles no solteras que en 1992 vivían en la Región del

Biobío. Los predictores finales que se mostraron como significativos en la modelización fueron la variable control representada por la edad de la mujer (siete cohortes); la educación de la mujer –utilizada como variable categórica con cinco estratos (sin educación, con educación básica incompleta, con educación básica completa, con educación secundaria incompleta, y al menos con educación secundaria completa); la ocupación de la mujer rural –que fue clasificada en cuatro categorías (dueñas de casa, trabajo en empleos remunerados, sin trabajo permanente y dependientes)–, y, finalmente, la religión, que fue categorizada en tres estratos (católica, evangélica y otras).

La edad resultó ser el factor que más variación de la variable dependiente explica en el nivel 1. La desviación causada por la edad fue de 26 613 puntos, lo que significa que la edad por sí sola explica 32.8% de la paridez marital en este nivel. En forma independiente, la educación reduce 12 161 puntos en paridez, lo que quiere decir que explica 15% de la variación de la variable dependiente. En este aspecto, es destacable que cuando la variable educación es agregada al modelo univariado de edad, la variación se reduce en 3 235 puntos adicionales, a un costo de 4 grados de libertad, lo que implica que en el nivel 1 el efecto neto de la educación es en términos reales de 4%. La ocupación da cuenta de otros 322.6 puntos de desviación, mientras que la religión agregó 151.7 puntos más de reducción, lo que en términos relativos significa 0.4% y 0.2% del total de la variación de la variable respuesta. Cuando todas las variables fueron modeladas conjuntamente, la desviación del modelo óptimo final fue de 30 301 puntos, lo que implica que todas las variables en su conjunto explican 37.3% de la variación individual de la paridez marital.

2. Variación de fecundidad entre distritos rurales (nivel 2)

En el nivel 2 de nuestra ecuación de niveles múltiples se modela el promedio de paridez total de los 348 distritos rurales como función de predictores exclusivamente de nivel 2 o distritales. Sin embargo, lo interesante de esta modelización de niveles múltiples es que los 348 distritos rurales están dentro de las 49 comunas, de modo que la variación de fecundidad que es estrictamente explicable a nivel de distrito debe por tanto modelarse como una parte diferente e independiente de aquella que es explicable en el nivel 3 o comunal. Para tales efectos, las variables a modelar son la edad, la educación, y el promedio de viviendas con

alcantarillado, que fueron las únicas variables pertinentes para explicar la paridez a nivel distrital. Pero la modelización no es simplemente de los promedios distritales de las variables, sino que abarca también las diferencias entre los respectivos promedios distritales y comunales, a fin de captar eficientemente en el modelo la parte que, en forma independiente de la comuna, debe ser atribuida estrictamente al distrito. La estimación estadística implicó entonces el ajuste de términos tales como $ddedu2$, variable centrada que expresa la diferencia entre el porcentaje de mujeres en el distrito con educación primaria incompleta y el promedio comunal de las mujeres con igual característica, esto es, $ddedu2 = medu2 - mmedu2$.

Este término ha sido denominado variación interdistrital pero intracomunal, para denotar la idea de que es una variación atribuible exclusivamente a los distritos, aunque considerando el hecho de que éstos forman parte de la jerarquía superior constituida por las comunas.

En el cuadro 8 puede observarse que el distrito explica 821.8 de los 81 661 puntos de la variación de la variable dependiente. Aunque esta magnitud no es muy grande, los 821.8 puntos de desviación atribuibles a los 347 grados de libertad rechazan la hipótesis nula de no existencia de efectos contextuales provenientes del distrito sobre la paridez de las mujeres rurales. De esa manera, los 821.8 puntos constituyen el máximo de variación que es posible explicar por medio de variables de nivel 2 o distritales. Según la selección de predictores enunciada en el cuadro 4, sólo los promedios de edad, de educación y de viviendas con alcantarillado, a nivel de distrito, reducen significativamente parte de esa variación total del nivel 2.

La modelización efectuada indica que a nivel distrital la variable más importante es el promedio de educación, que por sí sola explica 291.04 puntos, esto es, 35.4% de la variación de la paridez distrital. La edad reduce 104.9 puntos, con lo cual explica 12.8% de la variación de la variable dependiente en el nivel 2. Como se advierte, la educación es a nivel distrital casi tres veces más importante que el promedio de edad como predictor de la paridez. Cuando la variable edad es agregada al modelo de educación en el nivel 2, la explicación es de 45.2 puntos más, con una pérdida de 6 grados de libertad, lo que indica que el aporte de la edad al modelo es estadísticamente significativo. Ambas variables dan cuenta de 40.9% de la variación de la paridez a nivel distrital. El promedio de viviendas con alcantarillado agrega 1% de explicación de la variación en el modelo. Al agregar esta tercera variable al modelo de efectos distritales, se reduce 8.7 puntos más la desviación, con pérdida

de un grado de libertad, lo que es sumamente significativo y revela que el promedio de viviendas con alcantarillado es una variable contextual que debe ser agregada al modelo.

3. Variación de la paridez entre comunas (nivel 3)

La variación entre comunas, que en este estudio corresponde al nivel 3, modela la variación de la paridez promedio de las 49 comunas en términos de los predictores detectados como significativos en la etapa de selección. Las comunas explican 661 puntos de la variación de la variable dependiente, lo que corresponde a poco menos de 1% del total de la variación de la respuesta. Para los 48 grados de libertad envueltos, esta cantidad es muy significativa e implica que existe una porción de la variable respuesta que debe ser explicada por variables de nivel comunal.

En forma más pronunciada que en el nivel 2 o distrital, la variable promedio comunal de educación resultó el predictor más importante en el nivel 3, pues explica 50% de la variación de la paridez entre comunas. La edad contribuye en este nivel a explicar otro 23.5% de esa varianza. Sin embargo, el modelo aditivo de educación y edad explica 57.9%, dando cuenta así de 382.8 puntos de los 661 totales, lo que implica que la adición de la edad al modelo que abarca sólo la educación explica 54.4 puntos más de desviación, a un costo de 6 grados de libertad. En consecuencia, se justifica la incorporación de la variable edad como control en el modelo final de efectos comunales.

4. Efectos sobre la paridez detectados por la modelización de niveles múltiples

En esta sección se describen los efectos sobre la paridez provenientes de los predictores significativos seleccionados por el modelo de niveles múltiples utilizado en cada nivel de agregación, predictores que fueron sintetizados en el cuadro 8.

Lo que se pretende aquí es destacar principalmente los hallazgos referidos a las consecuencias contextuales de la paridez, más que profundizar en los modelos explicativos a nivel individual. La razón es simple: tal es el tema de nuestra investigación, y si bien hay décadas de fructífera investigación en lo referente al nivel individual, en la investigación contextual todo está aún en una etapa inicial.

Cuadro 9
COEFICIENTE DE REGRESIÓN DE POISSON Y VALORES *t*
PARA EL MODELO DE NÚMERO TOTAL DE HIJOS
DE MUJERES FÉRTILES DE LA REGIÓN DEL
BIOBÍO A NIVEL INDIVIDUAL

Variables	Coefficientes	Error estándar	Razón <i>t</i>
Intercepto	-0.093	0.003	-31.0
dedad2	0.278	0.021	13.2
dedad3	0.606	0.019	31.9
dedad4	0.887	0.019	46.7
dedad5	1.074	0.019	56.5
dedad6	1.208	0.019	63.6
dedad7	1.344	0.019	70.7
dedu2	-0.078	0.009	-8.7
dedu3	-0.180	0.007	-25.7
dedu4	-0.308	0.009	-34.2
dedu5	-0.494	0.011	-44.9
dcatol	-0.140	0.011	-12.7
devang	-0.058	0.012	-4.8
dtrabaj	-0.126	0.012	-10.5
dsintrab	-0.228	0.045	-5.1
dmdep	-0.051	0.019	-2.7

Fuente: Elaborado por el autor, sobre la base del Censo Nacional de Población y Vivienda de 1992.

Las variables edad, educación, religión y ocupación en el modelo intradistrital o de nivel individual han sido previamente centradas, esto es, utilizadas como diferencias individuales con respecto a sus medias distritales. Esto se expresó en la modelización utilizando, en el caso de edad, por ejemplo, los términos en la siguiente forma: $dedad2 = edad2 - edad2_{med}$. Los efectos de las variables significativas a nivel 1 o individual se observan en el cuadro 9.

Se estableció previamente que la variable edad fue incorporada en el modelo de niveles múltiples para captar los efectos de la exposición al riesgo de concebir de la mujer en ausencia de predictores asociados a la duración de la unión conyugal. En este sentido, es una variable control que se incluye en los modelos para ajustar los efectos de los restantes predictores individuales significativos.

Al interpretar los coeficientes de los modelos de regresión utilizados en esta sección, es conveniente hacer explícito su significado en los modelos de Poisson aquí ajustados. Cuando β_1 es el coeficiente de una variable indicador (0/1), por ejemplo, la exponenciación de este coeficiente $\exp[\beta_1]$ es el cambio proporcional en el número total de niños de la mujer cuando el valor de la variable cambia de 0 a 1, man-

teniendo constante todas las demás variables del modelo. En el modelo que se presenta en el cuadro 9, el coeficiente estimado de dedu2 es -0.078 , y de $\exp[-0.078]=0.92$. Por tanto, debemos concluir que las mujeres con educación primaria incompleta tienen en promedio 8% menos hijos que las mujeres sin educación, manteniendo la edad y los otros predictores de la ecuación constantes.

Como se muestra en el cuadro 9 los efectos individuales de la educación manteniendo constante la edad de las mujeres indican que el número de hijos de las mujeres se reduce significativamente a medida que aumenta el nivel educacional. La exponenciación de los coeficientes de regresión de Poisson indicados en el cuadro 9 deja ver, como se dijo recién, que las mujeres con educación primaria incompleta tienen 8% menos hijos que las mujeres sin educación. Esta reducción es de 17% para las mujeres con educación primaria completa; de 27% para las mujeres con alguna educación secundaria, y de 49% para las mujeres con al menos educación secundaria completa. Como indican los valores t de Student para cada coeficiente de Poisson de educación en el cuadro 9, esta reducción proporcional es significativa en cada estrato de educación. En conclusión, en el nivel 1 o individual, la educación tiene efectos claramente negativos sobre el número total de hijos de las mujeres rurales de la Región del Biobío.

La religión y la ocupación de la mujer también tienen efectos netos significativos a nivel individual. Mientras las mujeres pentecostales tienen 6% menos hijos que las mujeres protestantes y no creyentes, que constituyen la categoría de referencia, las católicas, en promedio, tienen 13% menos hijos que las de la categoría de referencia. En el caso de la ocupación, las mujeres sin trabajo tienen la menor tasa de fecundidad, con 20% de hijos menos que las mujeres dueñas de casa, que han sido tomadas aquí como categoría de referencia. Las mujeres con trabajo remunerado fuera del hogar tienen una fecundidad 12% menor que las dueñas de casa, mientras que las dependientes sólo tienen 6% de hijos menos que el estrato de referencia. Es claramente observable que la mayor educación, la religión católica y la cesantía son factores que reducen la fecundidad de las mujeres rurales en el nivel 1 o individual.

Al examinar los efectos de las variables predictoras significativas en el modelo conjunto con la variable control edad (véase el cuadro 10), puede observarse que los efectos contextuales de la educación van en similar dirección que los encontrados a nivel individual a partir del grupo con educación primaria completa (ddedu3).

El efecto neto que tiene el promedio de educación a nivel distrital se observa con bastante claridad en el cuadro 10. No se observan

Cuadro 10
COEFICIENTES DE REGRESIÓN DE POISSON Y VALORES *t*
PARA EL MODELO DE PARIDEZ TOTAL DE LAS
MUJERES RURALES EN EL NIVEL DISTRITAL

VARIABLES	COEFICIENTES	ERROR ESTÁNDAR	VALOR <i>t</i>
Intercepto	-0.001	0.0023	-0.43
ddedad2	-0.401	0.1668	-2.81
ddedad3	-0.477	0.1645	-2.90
ddedad4	-0.269	0.1593	-1.69
ddedad5	-0.158	0.1530	-1.03
ddedad6	-0.043	0.1677	-0.25
ddedad7	-0.088	0.1716	-0.51
ddedu2	0.094	0.1117	0.84
ddedu3	-0.383	0.0622	-6.15
ddedu4	-0.403	0.0617	-6.53
ddedu5	-0.920	0.0841	-10.94
ddalcant	0.163	0.0555	2.94

Fuente: Elaborado por el autor, sobre la base del Censo Nacional de Población y Vivienda de 1992.

diferencias significativas en la proporción de paridez que revelan las personas con educación primaria incompleta (ddedu2) y las personas sin educación (categoría de referencia en el modelo). Sin embargo, a partir de la categoría de las personas con educación primaria completa (ddedu3), puede observarse un efecto muy nítido sobre la paridez. El promedio de hijos de las personas con educación primaria completa es 33% inferior al promedio encontrado en las personas sin educación. El mismo porcentaje de reducción de la paridez promedio se da en las personas con educación secundaria incompleta (ddedu4). Sin embargo, el promedio de hijos de las personas en el distrito con al menos educación secundaria completa es 60% menor que el que muestran las personas sin educación.

El promedio de viviendas con alcantarillado en el distrito tiene un efecto positivo sobre la paridez distrital, lo que resulta muy diferente al sentido de la relación de orden cero, en que se había constatado una relación inversa entre esa característica y la paridez. Es probable que exista algún tipo de interacción estadística, no detectada, entre las variables alcantarillado y educación, lo que hace poco estable este efecto independiente.

Lo más importante de este hallazgo es que la educación tiene, simultáneamente, efectos individuales y contextuales sobre la paridez. Mejor dicho, la variable educación capta, a nivel distrital, una parte de la variación de la variable paridez de la mujer que no es atribuible al

Cuadro 11
COEFICIENTES DE REGRESIÓN DE POISSON Y VALORES *t*
PARA EL MODELO DE PARIDEZ TOTAL DE LAS
MUJERES RURALES A NIVEL COMUNAL

Variables	Coefficientes	Error estándar	Valor <i>t</i>
Intercepto	3.770	0.3073	12.3
mmedad2	-2.221	0.3886	-5.7
mmedad3	-2.144	0.3653	-5.9
mmedad4	-2.326	0.3677	-6.3
mmedad5	-2.125	0.3713	-5.7
mmedad6	-2.776	0.4167	-6.7
mmedad7	-1.500	0.3283	-4.6
mmedu2	-0.288	0.1965	-1.7
mmedu3	-0.577	0.1011	-5.7
mmedu4	-0.411	0.1023	-4.0
mmedu5	-1.208	0.1047	-11.5

Fuente: Elaborado por el autor, sobre la base del Censo Nacional de Población y Vivienda de 1992.

nivel individual. Éste es un efecto contextual propiamente tal, porque es una característica del distrito en cuanto agregado social, diferente a la característica individual expresada en la variable educación de cada mujer rural. Es importante destacar que, debido al uso de centramiento en ambos casos, no existe colinealidad entre la variable educación a nivel individual y la misma variable a nivel distrital.

En el cuadro 11 pueden observarse los coeficientes de regresión de Poisson a nivel comunal, en el modelo final de los efectos sobre la paridez provenientes del contexto comunal.

En el cuadro 11 se muestran los efectos comunales de los promedios de educación sobre los promedios de paridez, controlando por edad de las mujeres. Tampoco en el nivel comunal existen diferencias significativas entre el promedio de hijos de las mujeres con educación primaria incompleta y el de las mujeres sin educación, que forman el grupo de referencia. En cambio, las mujeres con educación primaria completa tienen en promedio 44% menos hijos que el grupo sin educación. Las mujeres con educación secundaria incompleta tienen sólo la tercera parte del promedio de hijos del grupo no educado. Finalmente, el grupo más educado en el nivel comunal reduce su número de hijos en 70% con respecto al grupo de mujeres sin educación. Como puede observarse, el efecto contextual de la educación en el nivel 3 es del mismo tipo que el observado en el nivel 2: a mayor promedio de educación comunal, menor es el número total de hijos.

CONCLUSIONES

El hallazgo principal de este estudio consistió en demostrar que la paridez en la Región del Biobío puede ser explicada mediante predictores que actúan a diferentes niveles. La organización jerárquica de la información censal permitió utilizar un modelo formado por tres niveles de análisis: 55 905 mujeres en edad fértil (nivel 1 o individual), que viven en 348 distritos rurales (nivel 2 o distrital), incluidos territorialmente en 49 comunas (nivel 3 o comunal).

Los datos del último Censo Nacional de Población y Vivienda, realizado en 1992, permitieron disponer de un conjunto de predictores de nivel individual para el número total de hijos declarados por las mujeres no solteras de la Región del Biobío. A estos predictores se agregaron datos censales referidos a la vivienda de las encuestadas y a algunas características socioeconómicas de sus hogares. Mediante el uso del Sistema de análisis estadístico (SAS) para la exploración de los archivos censales, de MLn para la organización en niveles múltiples de los archivos, y de Generalized Linear Interactive Modelling (GLIM) para la modelización final de los datos, se pudo determinar que la variación de la variable respuesta del estudio, el número total de hijos de las mujeres en edad fértil, puede ser explicada directamente, al menos en un 2%, mediante características descriptoras del contexto social en que viven las mujeres analizadas.

Aunque un 2% de explicación mediante factores contextuales puede parecer una proporción no importante, se debe subrayar que es una proporción estadísticamente significativa y constituye una porción de variación irreductible a la mera existencia de diferencias individuales. Tal porcentaje, que es explicable sólo en función de las diferencias de contexto social que rodean a la mujer rural, permite a la vez distinguir entre distritos y entre comunas. En la parte explicable a nivel de distrito se encontró que los promedios de educación y el promedio de viviendas con alcantarillado tienen efectos netos de la edad sobre la fecundidad. Los efectos de la educación son claramente inversos, lo cual significa que mientras mayor sea el grado de educación, menor será la fecundidad distrital. El promedio de viviendas con alcantarillado tiene efectos menos claros. Su relación bivariada con la paridez es negativa, mientras que su participación en el modelo final, aunque significativa, parece ocultar un efecto interactivo que no se intentó precisar en aras de la simplicidad de la interpretación.

A nivel comunal, sólo la educación tiene efectos contextuales sobre la paridez. Estos efectos son negativos y netos, variable que fue incluida debido a la necesidad de disponer en el modelo de una variable que permitiera captar la incidencia de la exposición al riesgo de concebir, en ausencia de datos censales relativos a la duración de la unión conyugal.

En resumen, merced a la naturaleza del modelo de estimación de efectos contextuales utilizado en el presente estudio, se logró determinar que sobre la paridez de la mujer fértil rural operan simultáneamente variables de nivel individual y de nivel contextual. Esta última dimensión fue obtenida mediante la utilización de dos niveles, el distrital y el comunal, que demostraron efectos significativos, independientes y simultáneos sobre la variable respuesta del estudio. Digno de la mayor atención resultó el hallazgo de que la educación fue el predictor simultáneo más significativo en los tres niveles del estudio. Esto demuestra que el uso exclusivo de esta variable a nivel individual, como se hace en la mayoría de las investigaciones tradicionales de un solo nivel, no capta la influencia de los factores educacionales que operan simultáneamente también en otros niveles superiores de agregación (en nuestro caso, el nivel distrital y el nivel comunal).

La utilización de variables centradas en el procedimiento de estimación estadística del método de niveles múltiples aseguró la inexistencia de colinealidad entre las mediciones de educación en los diferentes niveles, por lo que la detección de efectos significativos en los distintos contextos significa simplemente que la educación es un factor explicativo de la fecundidad en cada uno de los distintos niveles de agregación considerados. Al interpretar estos efectos se debería establecer que a mayor educación individual de las mujeres y a mayor grado de educación en el distrito rural y en la comuna, menor tiende a ser el número de hijos. En otras palabras, si las mujeres de más educación tienden a preferir un número reducido de hijos, tal efecto se acrecienta cuando viven además en una localidad en que están rodeadas de otras personas con alta educación, pues ese ambiente normativo tiene un peso extra que incidirá en una reducción aún más acentuada del número de hijos. Éste es el principal hallazgo del presente estudio, que ratifica que el análisis de niveles múltiples es una técnica promisoría, que será de gran provecho en la investigación de los efectos contextuales en las ciencias sociales en los próximos años.

BIBLIOGRAFÍA

- Agresti, Alan (1996), *An Introduction to Categorical Data Analysis*, Nueva York, John Wiley & Sons, Inc.
- Aitkin, Murray y otros (1989), *Statistical Modelling in GLIM*, Nueva York, Oxford University Press.
- Bilsborrow, Richard E. y David K. Guilkey (1987), "Community and Institutional Influence on Fertility: Analytical Issues", Population and Labor Policies Programme Working Paper, N° 157, Ginebra, Organización Internacional del Trabajo (OIT).
- Bilsborrow, Richard E. (1985), "Collecting community-level data for fertility analysis", *The Collection and Analysis of Community Data*, John Casterline (comp.), Voorburg, Países Bajos, Instituto Internacional de Estadística.
- Billy, John O.G. y otros (1989), *Final Report: Effects of Contextual Effects on Fertility Regulation and Fertility in the United States*, Seattle, Battelle Human Affairs Research Center, Universidad de Washington.
- Blalock, Hubert Jr. (1984), "Contextual-effects models: theoretical and methodological issues", *Annual Review of Sociology*, vol. 10.
- (1983), "Cross-level analyses", *The Collection and Analysis of Community Data*, John B. Casterline (comp.), Voorburg, Países Bajos, Instituto Internacional de Estadística.
- Boyd, Lawrence H. Jr. y Gudmund Iversen (1979), *Contextual Analysis: Concepts and Statistical Techniques*, Belmont, California, Wadsworth.
- Bryk, A.S. y S.W. Raudenbush (1988), *Hierarchical Linear Models: Applications and Data Analysis Methods*, Sage Publications.
- Casterline, John B. (comp) (1985), *The Collection and Analysis of Community Data*, Voorburg, Países Bajos, Instituto Internacional de Estadística.
- Chayovan, Napaporn B. (1982), "A Contextual Analysis of Demographic Phenomena in Rural Thailand", tesis para optar al grado de Ph.D., Departamento de Sociología, Universidad de Michigan.
- Entwisle, Barbara y otros (1986), "The multilevel dependence of contraceptive use on socioeconomic development and family planning program strength", *Demography*, N° 23.
- Espenshade, Thomas J. y Wenzhen Ye (1994), "Differential fertility within and ethnic minority: the effect of trying harder among Chinese-American women", *Social Problems*, vol. 41, N° 1.
- Gardner William, Edward Mulvey y Esther Shaw (1995), "Regression analysis of counts and rates: Poisson, Overdispersed Poisson, and Negative Binomial Models", *Psychological Bulletin*, vol. 118, N° 3.
- Gilchrist, Robert and Peter Green (1993), "The theory of Generalized Linear Models", *The GLIM System. Release 4 Manual*, Brian Francis y otros (comps.), Oxford, Clarendon Press.
- Goldstein, Harvey (1995), *Multilevel Statistical Methods*, segunda edición. Londres, Arnold.
- (1987), *Multilevel models in educational and social research*, Londres, Griffin.
- Hirschman, Charles y Philip Guest (1990), "Multilevel models of fertility. Determination in four Southeast Asian countries: 1970 and 1980", *Demography*, vol. 27, N° 3.
- Kass, Robert y Adrian Raftery (1995), "Bayes factors", *Journal of the American Statistical Association*, vol. 90, N° 430.

- Land, Kenneth C. y otros (1996), "A comparison of Poisson, negative binomial, and semiparametric mixed Poisson Regression Models", *Sociological Methods & Research*, vol. 24, N° 4.
- Lewis, Steven M. y Adrian Raftery (1995), "Comparing Explanations of Fertility Decline Using Event History Models with Unobserved Heterogeneity", *Technical Report*, N° 298, Seattle, Departamento de Estadísticas, Facultad de Estadísticas, Universidad del Estado de Washington.
- Liao, T. F. (1994), "Interpreting Probability Models: Logit, Probit, and Other Generalized Linear Models", Sage University Paper series on Quantitative Applications in the Social Sciences, N° 07-101, Thousands Oaks, California.
- Little, Roderick J.A. (1978), "Generalized Linear Models for cross-classified data from the WFS", *WFS Technical Bulletins*, N° 5, Voorburg, Países Bajos, Instituto Internacional de Estadística.
- Mason, W.M., G.Y. Wong y B. Entwisle (1983), "Contextual analysis through the Multilevel Linear Model", *Sociological Methodology 1983-84*, S. Leinhardt (comp.), San Francisco, Jossey Bass.
- McCullagh, P. y J.A. Nelder (1983), *Generalized Linear Models*, Cambridge, University Press.
- Merino, José Manuel (1993), "Contextual Effects on Current Use of Modern Contraceptive Methods: Service Availability of Family Planning and Contraceptive Prevalence in Rural Colombia", tesis para optar al grado de Ph.D., Austin, Departamento de Sociología, Universidad de Texas.
- Nizamuddin, M. (1979), "The Impact of Community and Program Factors on the Fertility Behavior of Rural Pakistani Women", tesis para optar al grado de Ph.D., Departamento de Planificación Demográfica, Escuela de Salud Pública, Universidad de Michigan.
- Pullum, Thomas W. y José M. Merino Escobar (1994), "Multi-level methods to estimate the impact of community services on fertility, contraceptive use, and child health. XIII World Conference of Sociology. International Sociological Association (ISA) (Bielefeld, Germany)", *Sociological Abstracts*, N° 173, julio.
- Pullum, Thomas W. (1995), "The Multivariate Analysis of Recent Fertility", Working Paper, Austin, Centro de Investigación Demográfica, Universidad de Texas.
- (1991a), "The Relationships of Service Availability to Contraceptive Use in Rural Guatemala", DHS Working Papers, N° 2. Columbia, Maryland, Instituto para el Desarrollo de Recursos (IRD)/Macro International.
- (1991b), "Community-Level Data Collection and Analysis", *Demographic and Health Survey Conference*, Washington, D.C.
- (1989a), "Models for multi-level analysis using variation between and within levels", *Final Report: Effects of Contextual Effects on Fertility Regulation and Fertility in the United States*, Seattle, Battelle Human Affairs Research Center, Universidad de Washington.
- (1989b), "The timing of first birth in the United States: a multi-level analysis using the 1982 National Survey of Family Growth", *Final Report: Effects of Contextual Effects on Fertility Regulation and Fertility in the United States*, Seattle, Battelle Human Affairs Research Center, Universidad de Washington.
- (1987), "Analytical Methodology", *The World Fertility Survey. An Assessment*, John Cleland y Chris Scott (comps.), Cambridge, Oxford University Press.
- Pullum, Thomas W. y otros (1984), "Effects of Contextual Factors on Fertility Regulation and on Fertility", Technical Proposal, NICHD RFP-DBS-84-14, inédito.

- Raftery, Adrian y otros (1993), "Event History Modeling of World Fertility Survey Data", Working Paper, N° 93-1, Centro de Estudios de Demografía y Ecología, Universidad del Estado de Washington.
- Raftery, Adrian (1994), "Bayesian model selection in social research", *Sociological Methodology*, Cambridge, Massachusetts.
- (1993), "Demand or Ideation? Evidence from the Iranian Marital Fertility Decline", Working Paper, N° 94-1, Centro de Estudios de Demografía y Ecología, Universidad del Estado de Washington.
- (1988), "Approximate Bayes factors for generalized linear models", Technical Report, N° 121, Departamento de Estadísticas, Facultad de Estadísticas, Universidad del Estado de Washington.
- (1986), "Choosing models for cross-classifications", *American Sociological Review*, N° 51.
- Rasbash, Jon y Geoff Woodhouse (1996), *MLN Command Reference*, version 1.0a, University of London.
- Tsui, Amy Ong (1985), "Community effects on contraceptive use", *The Collection and Analysis of Community Data*, John B. Casterline (comp.), Voorburg, Países Bajos, Instituto Internacional de Estadística.
- Woodhouse, Geoff y otros (1996), *Multilevel Modelling Applications: A Guide for users of MLN*, Londres, University of London.