

TASA DE NATALIDAD Y VARIABLES
SOCIO-ECONOMICAS:
UNA NOTA METODOLOGICA

Fernando Cortés
(Facultad Latinoamericana
de Ciencias Sociales, FLACSO)

Angel Flisfisch
(Facultad Latinoamericana
de Ciencias Sociales, FLACSO)

NATALITY RATE AND SOCIO-ECONOMIC VARIABLES:
A METHODOLOGICAL ACCOUNT

SUMMARY

There is strong empirical evidence pointing to the existence of relationships among several social and economic variables and the birth rate, as shown in an intercorrelation matrix for Latin American countries. The need to employ techniques of analysis of a more sophisticated and powerful nature, leads to the consideration of regression analysis as a suitable tool for the detection of more structured empirical relations.

Given the features presented by the intercorrelation matrix, the rank condition required by regression analysis is not fulfilled and, as a previous stage to it, it is necessary to summarize the data in an adequate form. This is done by means of principal component analysis. We have been able to discern two factors accounting for 86.7 per cent of the total variance of the independent variables. Employing the usual criteria for factor interpretation, the first is baptized as socio-economic growth and the second as socio-economic heterogeneity. Both factors explain around 81 per cent of the birth rate variance. The complexities involved in the interpretation of the second factor, leads to the consideration of some interaction models relating the birth rate to selected economic and social variables.

Pocas personas discutirían la validez general de aquellas aseveraciones que relacionan algunos aspectos esenciales de la dinámica social (tales como procesos de crecimiento económico, urbanización, desarrollo educacional) con el movimiento de ciertas variables demográficas.

Pensamos, fundamentalmente, en variables clásicas como, por ejemplo, natalidad y mortalidad.^{1/}

Usualmente, se considera que proposiciones del tipo de: "a mayor crecimiento económico menor tasa de natalidad", "a mayor nivel educacional menor tasa de natalidad", "a mayor población urbana menor tasa de natalidad", pertenecen por derecho propio al discurso demográfico contemporáneo. Ciertamente es que las actuales tendencias de dicho pensamiento apuntan hacia explicaciones e intentos de análisis que suponen una alta complejidad en las relaciones mutuas entre variables. Una posible actitud, fundamentada en esta nueva orientación teórica, es la de, simplemente, olvidar el nivel de análisis que implica el tipo de proposiciones recién enunciadas, y rechazarlas por constituir generalizaciones relativamente carentes de contenido.

Por el contrario, en este trabajo intentamos demostrar que esas aseveraciones pueden ser vistas como puntos de partida útiles en el intento de comenzar a recuperar la complejidad que caracteriza la realidad social.

En verdad, la forma adoptada la mayoría de las veces por la investigación empírica que fundamenta las relaciones establecidas, es de naturaleza tal que, por una parte, facilita bastante el juicio negativo y el rechazo consiguiente de estas generalizaciones y, de otra, hace imposible la introducción de una mayor complejidad en el análisis de datos y en las conclusiones derivadas.

Normalmente, se inicia el análisis empírico a partir de una matriz de intercorrelaciones. (Véase el cuadro 1).

La simple inspección de esta matriz permite constatar, en principio, la validez de las proposiciones en cuestión.

Si decidiéramos detener el análisis en este punto, dándonos por satisfechos con los hallazgos encontrados, estaríamos suponiendo que los efectos de las distintas variables asumen el carácter de independientes entre sí, es decir, cada una de ellas actuaría con un efecto propio y por separado sobre la variable a explicar. Sin embargo, la misma matriz de intercorrelaciones nos enseña que ésta no es la situación. Muy por el contrario, ella muestra la existencia de toda una estructura de relaciones al interior de las variables explicativas y de ellas con la variable explicada.

Por otra parte, el uso de un marco teórico provisto de una mínima sofisticación debería también hacernos concluir que un procedimiento de tal naturaleza es profundamente inadecuado como instrumento analítico.

En esencia, mientras permanezcamos a este nivel técnico, la situación se caracterizaría por una significativa falta de adecuación entre el

^{1/} Para una revisión cuidadosa y detallada de esta bibliografía véase Raúl Urzúa, *La Investigación en Ciencias Sociales y las Políticas de Población: Sugerencias para la Programación de Actividades de PISPAL*, Unidad Central PISPAL, abril de 1974, Santiago de Chile.

discurso teórico y el instrumento de análisis: se intenta capturar una realidad que se postula como compleja, por medio de técnicas que suponen esa realidad como simple. En presencia de este problema, usualmente se recurre ya sea al análisis de correlación parcial o al de regresión.

El primero de ellos presenta algunas limitaciones que con frecuencia pasan desapercibidas. En primer lugar, existe una dificultad de orden práctico. Por lo común, el análisis de correlación utiliza sólo correlaciones parciales de primer orden; dada la lógica de esta técnica, este modo de utilizarla implica suponer necesariamente que las variables explicativas contempladas no presentan relaciones significativas con la variable "controlada". Ciertamente que esta limitación puede subsanarse mediante el recurso de correlaciones parciales de orden superior. Si el número de variables es relativamente elevado, el orden en el cual hay que detenerse se aumenta correspondientemente. Pero aun así subsiste una segunda dificultad. En efecto, los modelos de correlaciones

Cuadro 1

MATRIZ DE INTERCORRELACIONES ENTRE
VARIABLES SOCIO-ECONOMICAS

	Correlaciones entre variables explicativas						Y_N
	X_1	X_2	X_3	X_4	X_5	X_6	
X_1	1,000	-	-	-	-	-	-
X_2	-0,739	1,000	-	-	-	-	-
X_3	0,835	-0,692	1,000	-	-	-	-
X_4	0,504	-0,695	0,538	1,000	-	-	-
X_5	0,701	-0,673	0,863	0,700	1,000	-	-
X_6	0,811	-0,711	0,700	0,756	0,849	1,000	-
Y_N	-0,725	0,459	-0,851	-0,416	-0,829	-0,724	1,000

en que:

X_1 : Producto interno bruto por habitante (en US\$ a precios de 1960).

X_2 : Producto agrícola bruto como porcentaje del producto interno bruto.

X_3 : Porcentaje de población urbana.

X_4 : Matrícula primaria como porcentaje de la población de 7 a 13 años.

X_5 : Matrícula secundaria como porcentaje de la población de 15 a 19 años.

X_6 : Alfabetizados de 15 años y más.

Y_N : Tasa de natalidad por mil habitantes. ^{2/}

^{2/} Los valores asociados a cada una de estas variables, así como las fuentes, se encuentran en la matriz de datos, anexa a este trabajo.

parciales son determinísticos, es decir, es necesario suponer que la variación en la variable a explicar se encuentra determinada en su totalidad por la variación en las variables explicativas. Este supuesto es bastante fuerte y, en general, nuestras nociones acerca de la realidad social nos llevarían a rechazarlo, lo cual no agota las limitaciones de esta técnica, por cuanto el modelo subyacente a la utilización de las correlaciones parciales supone linealidad estricta entre las variables. Además, manteniéndonos aún a nivel meramente técnico, hay que destacar que, en sentido estricto, el modelo de correlación supone distribución normal conjunta de las variables. Esta condición agrega otra fuerte restricción a su utilización.

Finalmente, hay que destacar que la utilización de correlaciones parciales presenta dificultades derivadas de su pobreza analítica. La única interpretación posible de un coeficiente de correlación es en términos de porcentaje de varianza explicada. En el caso específico de los parciales, ello significa que se trata de ese porcentaje, pero una vez que se ha eliminado el impacto lineal de un conjunto de variables.

En cambio, el modelo de regresión nos permite no sólo una medición adecuada del porcentaje de varianza explicada, sino que además posibilita la individualización de los efectos atribuibles a cada variable explicativa, y ello en términos tales que se abre una rica gama de posibilidades de interpretación a nivel de un lenguaje más natural. Así, por ejemplo, es usual la utilización en el lenguaje preteórico de nociones tales como: tasas de cambios; relaciones entre ellas; sus variaciones; impactos marginales; dirección de los efectos. Todas estas nociones pueden traducirse adecuadamente por medio de los coeficientes de regresión, lo que no acontece con las correlaciones parciales.

Esta sola característica hace que el modelo de regresión sea analíticamente más potente que el de correlación, pero además elimina las otras dificultades que hemos constatado. Sin embargo, hay que advertir que el modelo de regresión múltiple supone, entre otros, que las variables explicativas son relativamente independientes. En los casos en que ello no ocurre se deben encarar complicados problemas de estimación.

El asunto que nos ha ocupado hasta ahora dice relación con las técnicas adecuadas para establecer vinculaciones entre variables. Un problema algo diferente es el que se plantea cuando tenemos un conjunto de variables relacionadas entre sí, y el objetivo reside en capturar y expresar de manera sintética la estructura de relaciones subyacentes. La matriz de intercorrelaciones que hemos presentado suscita precisamente este problema, según lo hemos indicado con anterioridad.

Existe un conjunto de técnicas alternativas que permiten dar una respuesta a esta cuestión. Desafortunadamente, y a diferencia de lo que acontece con la evaluación de los modelos de regresión y correlación,

no disponemos de buenos criterios que nos permitan justificar adecuadamente la elección de una técnica en lugar de otras.

Para los fines de este trabajo, hemos recurrido al así denominado análisis de componentes principales. ^{3/}

Nuestro propósito reside entonces en el intento de expresar la complejidad de las relaciones que hemos advertido en la matriz de intercorrelaciones. Si tenemos éxito, ello quizás pueda permitirnos ofrecer una explicación del movimiento de la tasa de natalidad en términos algo más razonables que la pura constatación de las correlaciones por pares de variables.

El análisis de componente principal de la matriz de datos, ^{4/} nos permite distinguir dos factores, que conjuntamente explican alrededor de un 87 por ciento de la varianza total, es decir, de la varianza atribuible a nuestras seis variables explicativas. (Véase el cuadro 2).

Cuadro 2

FACTORES RESULTANTES DEL ANALISIS DE COMPONENTES
PRINCIPALES Y SUS CONTRIBUCIONES A LA VARIANZA
TOTAL

Factores	1	2	3	4	5	6
Contribución	0,767	0,100	0,064	0,045	0,023	0,003
Contribución acumulada	0,767	0,867	0,931	0,976	0,999	1,000

^{3/} La exposición del método de componentes principales se encuentra en una gran cantidad de textos. Entre ellos destacamos:

Johnston, J., *Econometrics Methods* (segunda edición), McGraw-Hill Book Company, Nueva York, 1972, págs. 322-330.

Malinvaud, E., *Métodos Estadísticos de la Econometría*, Ediciones Ariel, Barcelona, 1967, págs. 682 y 683.

Theil, Henry, *Principles of Econometrics*, John Wiley and Sons Inc., Nueva York, 1971, págs. 46-55.

^{4/} El principio básico sobre el cual está construido el método de componentes principales, es el de maximizar ordenadamente el poder explicativo de los factores. De este modo, para obtener el primer factor se busca aquella combinación lineal de las variables, que dé cuenta de la mayor parte posible de la varianza. Este procedimiento es aplicado para obtener el segundo factor, pero esta vez, se impone como condición de que él no se encuentre correlacionado con el primero (supuesto de ortogonalidad). Del tercero al *n*-ésimo factor se opera tal cual que para el segundo, pero considerando que la ausencia de correlación es válida para cualquier par de factores.

La varianza total de los factores está construida a base de una combinación lineal de variables, que nosotros hemos tipificado con el propósito de eliminar problemas que pudiesen suscitarse por las unidades de medida. Tenemos variables expresadas en dólares y en porcentajes.

El cuadro 1 presenta la contribución a la varianza total de cada uno de los seis factores. Sólo hemos conservado los dos primeros, en razón de que la inclusión del tercero tiene un efecto sensiblemente menor en términos de varianza explicada. El mismo argumento nos lleva a desechar los factores restantes. Obviamente, esta decisión es relativamente arbitraria y constituye una de las debilidades del análisis de componentes principales, puesto que no existen criterios precisos para conservar o eliminar factores.

Una característica que es común a toda esta clase de técnicas radica en que ellas están construidas de manera tal que garantizan la obtención de resultados. En consecuencia, estos instrumentos no contemplan la posibilidad de discriminar entre situaciones caracterizadas por la existencia de factores y aquellas en que éstos no existen.

Lo anterior implica dificultades de interpretación bastante serias. Sin embargo, y no obstante lo que comúnmente se cree, existen criterios para otorgar significado a las dimensiones básicas seleccionadas.

Un primer criterio consiste en examinar los pesos de cada variable al interior de cada factor. En efecto, un factor cualquiera es simplemente una combinación lineal de las variables consideradas. De este modo, los pesos asignados a cada una de las variables pueden ser utilizados para investigar la importancia relativa de cada una de ellas.

Las expresiones siguientes nos muestran la forma en que se descompone cada factor:

$$F_1 = 0,454 X_1' - 0,443 X_2' + 0,456 X_3' + 0,408 X_4' + 0,473 X_5' + 0,476 X_6' \quad (1)$$

$$F_2 = -0,206 X_1' - 0,490 X_2' - 0,470 X_3' + 0,911 X_4' - 0,665 X_5' + 0,108 X_6' \quad (2)$$

en que F_1 y F_2 simbolizan el primer y segundo factor, respectivamente.

La ecuación (1) constituye algo así como una media aritmética simple, puesto que los pesos de las distintas variables son muy similares. La única dificultad aparente podría estar dada por el signo negativo que afecta al peso de la segunda variable. Este pequeño enigma se despeja de inmediato si se recuerda que esa variable es producto agrícola (expresado como porcentaje del producto interno bruto). Basta con tomar su complemento ($1 - X_2$: participación en el producto bruto interno de los sectores no agrícolas) para que el signo que afecta al peso se torne positivo.

A primera vista, el que el primer factor constituya un simple promedio aritmético puede parecer altamente desilusionante. Sin embargo, si recordamos que la matriz de correlaciones nos muestra fuertes relaciones entre *todas* las variables, esta característica del factor no debe producir extrañeza. Para que el análisis entregue factores que privilegien ciertas variables es necesario que la matriz de

intercorrelaciones presente discontinuidades significativas. Puesto que éste no es el caso de nuestra matriz, el primer factor está revelando una característica esencial de la estructura subyacente en los datos.

Considerando la naturaleza de F_1 , hemos decidido bautizarlo como *crecimiento socio-económico*. Según se verá más adelante, este factor distribuye las unidades nacionales en términos compatibles con nuestras nociones e intuiciones acerca de lo que sustantivamente significaría una dimensión como la indicada.

En el caso del segundo factor, el auxilio que nos presta la ecuación (2) no es tan claro. Dado el alto peso que tiene la variable X_4 (matrícula primaria como porcentaje de la población de 7 a 13 años), uno tendería a interpretarlo en términos de una dimensión que refleja el nivel educacional elemental alcanzado por una unidad nacional. Esta conclusión podría considerarse como de mayor plausibilidad si se examina el cuadro 3.

Cuadro 3

COEFICIENTES DE CORRELACION ENTRE LAS VARIABLES
Y LOS FACTORES

	X_1	X_2	X_3	X_4	X_5	X_6
F_1	0,870	-0,850	0,910	0,788	0,910	0,946
F_2	-0,178	-0,310	-0,313	0,526	-0,210	0,000

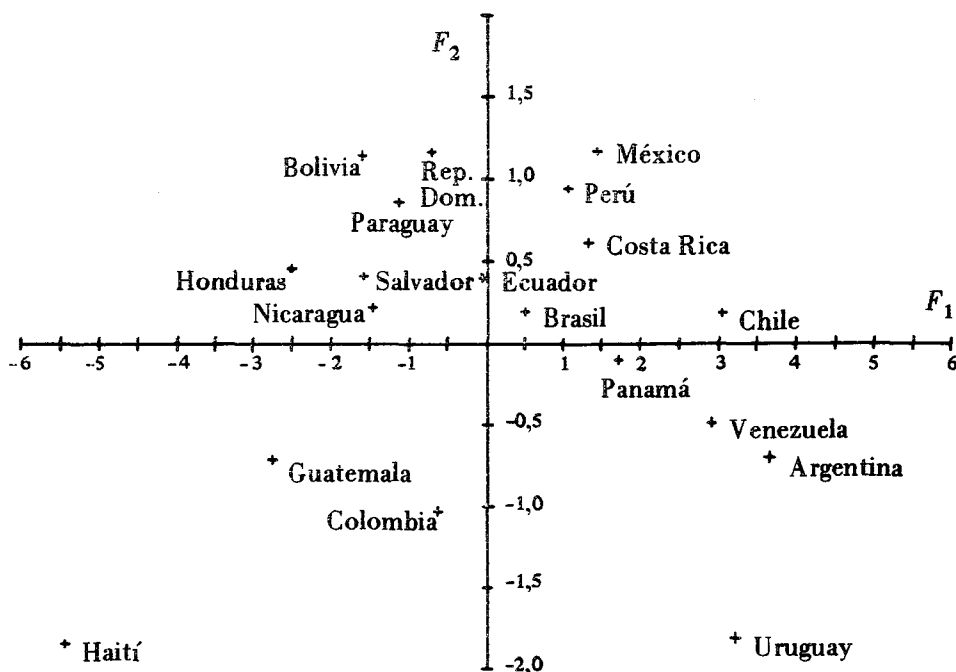
En efecto, la única variable que muestra una relación relativamente alta con F_2 es precisamente X_4 . De paso podemos señalar cómo este segundo criterio utilizado para la interpretación de los factores apoya nuestra interpretación asignada a F_1 . Sin embargo, si retornamos a la ecuación (2), observamos que esa atribución de significado no es tan clara: por una parte, la variable X_5 (matrícula secundaria como porcentaje de la población de 15 a 19 años) tiene un peso bastante alto y de signo contrario al de X_4 ; además, resulta que la contribución a F_2 de las dos variables económicas (X_1 y X_2), conjuntamente con la contribución de la variable X_3 (porcentaje de población urbana), son de naturaleza tal que tienden a contrapesar la contribución del nivel elemental de educación.

En la práctica se utiliza un tercer criterio para interpretar los factores. Éste consiste en examinar las posiciones que toman las distintas unidades de análisis en el espacio definido por los factores o dimensiones. Estas posiciones se muestran en el gráfico 1.

Si centramos la atención exclusivamente en el eje vertical, nos daremos cuenta de que la interpretación dada a F_2 (educación

Gráfico 1

POSICIONES DE LOS PAISES EN EL ESPACIO BIDIMENSIONAL
DEFINIDO POR LOS FACTORES "CRECIMIENTO SOCIO-ECONOMICO" (F_1)
Y "HETEROGENEIDAD SOCIO-ECONOMICA" (F_2)



elemental) es realmente absurda: en el extremo negativo encontramos dos países con posiciones casi idénticas: Haití y el Uruguay. Acontece que el Uruguay presenta niveles educacionales muy altos y Haití, por el contrario, es el país de más bajo nivel. Esta sola evidencia debería llevarnos a rechazar la interpretación postulada. Quizás convenga entonces preguntarse acerca de qué es lo que tienen en común estas dos unidades nacionales. La característica que ambas comparten es la de presentar valores relativamente altos en *todas* las variables (el caso del Uruguay) o bien valores bajos en las mismas (el caso de Haití). Examinando nuevamente el gráfico, nos percatamos, a medida que abandonamos el extremo negativo de la dimensión, que esta característica va desapareciendo. Así, encontramos en el extremo positivo que los países allí ubicados se caracterizan por claras discontinuidades o "desequilibrios" entre los valores del conjunto de variables.

En el caso de Bolivia se tiene una unidad nacional caracterizada por valores bajos en todas las variables, salvo X_4 , que presenta un valor alto y "no correspondiente" con los restantes puntajes. Para República Dominicana acontece exactamente lo mismo. Además, los dos países se caracterizan por una estructura agraria que tampoco "corresponde" a su

situación general: la participación del sector agrícola en el total de la economía (X_2) es más baja que la que presentan otros países en situaciones similares.

Un examen detallado de la matriz de datos ^{5/} nos indica que esta dimensión puede interpretarse en términos de estas consideraciones. Es decir, el segundo factor constituye una medida que refleja precisamente "no correspondencia" entre los valores de las variables para una unidad nacional cualquiera. Si bien es cierto que los indicadores del sistema educacional juegan un papel preponderante en la existencia de esta clase de discontinuidades, no lo es menos que también la dimensión en cuestión refleja quiebres menos evidentes y no atribuibles a las variables educacionales. En consecuencia, hemos optado por designar a F_2 como *heterogeneidad socio-económica*.

Claramente, existen nociones más sugestivas y elegantes que podrían utilizarse para etiquetar a F_2 . La primera que se viene a la mente es la de heterogeneidad estructural. Este concepto contiene una carga teórica bastante fuerte; parecería que, de manera necesaria, él implica o connota una teoría de estadios de desarrollo que constituiría un patrón para juzgar la adecuación, o su ausencia, del estado de una unidad nacional en relación con ciertas secuencias postuladas como típicas.

Con el propósito de evitar un compromiso de esa naturaleza, hemos preferido utilizar una noción más neutra. En definitiva, creemos que la dimensión rescatada por el segundo factor puede mirarse como reflejo de una característica de la estructura global de los datos, la que es susceptible de interpretaciones teóricas diversas.

Una vez resuelto nuestro primer problema, es decir, expresada de manera sintética la estructura de relaciones entre las variables explicativas, podemos preguntarnos acerca del efecto de esta estructura sobre la tasa de natalidad.

Para ello hemos efectuado un análisis de regresión de F_1 y F_2 sobre la tasa de natalidad (Y_N). Los puntajes que asumen los distintos países en los factores y los correspondientes valores de la tasa de natalidad se presenta en el cuadro 4.

La regresión sobre estos datos arroja el siguiente resultado:

$$Y_N = 38,392 - 2,630 F_1 + 4,129 F_2 \quad (3)$$

(0,867) (0,377) (0,953)

$$R^2 = 0,807 \quad R = 0,898$$

^{5/} Esta matriz se encuentra anexa.

A juzgar por el valor del coeficiente de determinación, podemos afirmar que el ajuste es bastante bueno: ambos factores explican cerca de un 81 por ciento de la variabilidad en la tasa de natalidad. El signo del coeficiente de regresión asociado a F_1 calza con las ideas previas respecto a la relación entre crecimiento socio-económico y tasa de natalidad. El coeficiente asociado a F_2 nos dice que a mayor heterogeneidad socio-económica, mayor tasa de natalidad. ^{6/} Puesto que este resultado no es en absoluto obvio, conviene preguntarse acerca de la magnitud del impacto *propio* de F_2 , ya que podría suceder que F_1 explicara casi en su totalidad el valor del coeficiente de determinación.

La ecuación (4) nos indica que éste no es el caso:

$$Y_N = 38,565 - 2,585 F_1$$

$$R^2 = 0,581 \quad R = 0,762 \quad (4)$$

Cuadro 4

CRECIMIENTO SOCIO-ECONOMICO, HETEROGENEIDAD
SOCIO-ECONOMICA Y TASA BRUTA DE NATALIDAD
POR PAISES

Países	F_1	F_2	Y_N
Argentina	3,56	-0,71	21,80
Venezuela	2,85	-0,50	36,06
México	1,37	1,16	42,00
Panamá	1,66	-0,12	36,08
Uruguay	3,15	-1,82	20,83
Chile	3,01	0,17	25,88
Costa Rica	1,27	0,60	33,44
Brasil	0,45	0,18	37,12
Perú	1,00	0,93	41,02
Colombia	-0,68	-1,05	40,61
Ecuador	-0,08	0,38	41,76
Guatemala	-2,81	-0,73	42,82
Nicaragua	-1,52	0,20	48,32
El Salvador	-1,63	0,39	42,17
Paraguay	-1,19	0,85	42,25
Rep. Dominicana	-0,78	1,15	45,84
Honduras	-2,56	0,44	49,26
Bolivia	-1,66	1,13	43,72
Haití	-5,50	-1,86	41,98

^{6/} No debe olvidarse que los factores han sido construidos con variables tipificadas.

En realidad, al introducir F_2 en el modelo conseguimos explicar el 53 por ciento de la varianza no explicada por F_1 . ^{7/} Este resultado es suficiente para conservar el modelo expresado por la ecuación (3).

El problema que se suscita ahora es de índole más bien sustantiva. Una relación entre variables establecida empíricamente carece de interés si no es posible atribuirle algún sentido o significado plausible, y lo cierto es que esta atribución de sentido se torna bastante difícil si se considera que el segundo factor expresa una característica compleja de la globalidad de los datos.

El análisis realizado es útil en cuanto nos indica que existen relaciones no simples entre las variables en cuestión, pero es incapaz por sí mismo de identificar de manera específica cuáles y de qué naturaleza son esas relaciones.

Sabemos que la heterogeneidad socio-económica afecta la tasa de natalidad, pero lo que ahora interesa es examinar en detalle en qué consistiría y cómo se relaciona con la variable explicada.

Nuestro punto de partida está dado por la relación entre heterogeneidad socio-económica y tasa de natalidad. Con el propósito de comprender mejor su significado, volvamos al gráfico 1 y concentrémonos en el primer cuadrante. El conjunto de países allí incluidos presentan un grado de crecimiento socio-económico relativamente más alto y, en consecuencia, tasas de natalidad relativamente más bajas; sin embargo, la circunstancia de que ellos se caracterizan por un grado de heterogeneidad más alto implica que esta expectativa no se satisfará totalmente. Inversamente, en el caso de aquellos países que se encuentran en el cuarto cuadrante, esa influencia positiva de la heterogeneidad socio-económica se encuentra bastante más disminuida, es decir, en estos casos el impacto del crecimiento socio-económico se produce en plenitud, precisamente por la ausencia de la *condición* cuya presencia caracterizaba los casos anteriores. Podemos analizar de manera similar las observaciones ubicadas en el segundo y tercer cuadrante.

^{7/} El 53 por ciento se obtuvo por medio de:

$$\frac{R^2 Y_{N, F_1, F_2} - R^2 Y_{N, F_1}}{1 - R^2 Y_{N, F_1}}$$

en que:

$R^2 Y_{N, F_1}$ es el coeficiente de determinación entre la tasa de natalidad y el primer factor; y

$R^2 Y_{N, F_1, F_2}$ es el coeficiente de determinación de esa tasa cuando se usan como variables explicativas ambos factores.

El denominador expresa el porcentaje de varianza no explicado por el factor 1 y el numerador, la ganancia obtenida en la explicación de dicha varianza por el hecho de incluir el segundo factor. Por lo tanto, la razón expresa el porcentaje de varianza explicada por el segundo factor respecto al total *posible* de explicar.

La proposición que derivamos de este examen es que, si bien existe un efecto del conjunto de las variables socio-económicas, impacto cuyo sentido es el usualmente esperado, ellas operan condicionadas por el grado de heterogeneidad presente.

En términos más específicos, esta interpretación nos estaría diciendo, por ejemplo, lo siguiente: el grado de crecimiento económico (en sentido estricto) tiende a hacer descender la tasa de natalidad, pero su efecto varía en relación con los valores adoptados por otras variables. Vale decir, el modelo o los modelos adecuados para capturar la relación entre una variable socio-económica y el comportamiento reproductivo de la población, deben tomar en cuenta que el efecto de una variable específica no es independiente de los niveles alcanzados por las otras.

Sobre la base de esta afirmación, procederemos a investigar los efectos de algunas variables escogidas, representativas de cada uno de los sectores o aspectos de la sociedad expresados en nuestros datos: economía, población urbana y educación. Para el aspecto económico, hemos escogido X_1 ya que es el indicador comúnmente utilizado como medida de nivel de crecimiento económico. Respecto al sector educacional, hemos escogido X_4 (Instrucción primaria), puesto que esta variable desempeña una función central en la existencia de discontinuidades o heterogeneidades.

El problema reside entonces en dar una expresión formal adecuada a la noción recién desarrollada, conocida en el lenguaje técnico como interacción entre variables. Supongamos que tenemos dos variables explicativas: Z y W y una variable a explicar Y . Aceptemos, además, que el efecto marginal de Z sobre Y depende del nivel de W , y viceversa. Esta idea se puede expresar mediante la ecuación:

$$Y = aZW$$

donde a es constante.

Al tomar derivadas parciales de esta expresión, se puede apreciar que el impacto marginal de cada una de las variables independientes depende del valor absoluto de la otra:

$$\frac{\partial y}{\partial z} = aw \qquad \frac{\partial y}{\partial w} = az$$

En concordancia con lo anterior, hemos postulado el modelo:

$$Y_N = a_0 + a_1 X_1 X_4 + a_2 X_3 X_4 + e_1 \qquad (5)$$

Con el objeto de tener un criterio que nos permita evaluar la ganancia obtenida mediante el uso de este modelo, en comparación con formas

más simples de análisis, hemos construido también uno agregativo simple para las mismas variables:

$$Y_N^i = b_0 + b_1 X_1 + b_2 X_3 + b_3 X_4 + e_2 \quad (6)$$

Los resultados de los ajustes de estas ecuaciones se muestran en el cuadro 5.

Cuadro 5
RESULTADOS DEL AJUSTE DE LOS MODELOS (5) Y (6)

Modelo	
(5)	$Y_N = 50,634 + 0,000 X_1 X_4 - 0,004 X_3 X_4$
	$R^2 = 0,773 \quad R_C^2 = 0,745$
(6)	$Y_N = 50,729 - 0,002 X_1 - 0,390 X_3 + 0,027 X_4$
	(4,810) (0,008) (0,114) (0,068)
	$R^2 = 0,720 \quad R_C^2 = 0,576$

La simple inspección de los coeficientes de determinación, una vez corregidos según los distintos grados de libertad (R_C^2), nos permite optar por el modelo de la ecuación (5).

Aun así, el modelo correspondiente a la ecuación (6) arroja un resultado de interés: el coeficiente asociado a X_1 (producto interno bruto per cápita) es no significativo. Este resultado se ve corroborado por cuanto en el otro modelo el coeficiente que afecta a la variable en que el ingreso per cápita entra en la interacción, es simplemente igual a cero.

Este resultado no deja de ser sorprendente. Con el propósito de explorar el posible impacto de la variable X_1 , hemos planteado un tercer modelo, suponiendo ahora que sólo existe interacción entre urbanización y educación básica, mientras que el nivel de crecimiento económico opera independientemente. (Véase el cuadro 6).

Como se puede apreciar, el impacto del ingreso es nuevamente mínimo y seguramente no significativo.

La idea subyacente en la construcción del modelo expresado por la ecuación (5) es que los efectos, tanto de la variable económica como de la variable urbana, dependen del nivel educacional alcanzado. En otras palabras, si bien aceptamos que el comportamiento reproductivo encuentra determinaciones por la economía, ellas sólo entrarían a

Cuadro 6

RESULTADO DEL AJUSTE DEL MODELO EN QUE EL
INGRESO PER CAPITA OPERA DE MANERA ADITIVA,
Y LA EDUCACION BASICA Y LA URBANIZACION DE
MODO INTERACTIVO

$$Y_N = 50,687 + 0,001 X_1 - 0,004 X_3 X_4$$

$$R^2 = 0,783$$

$$R_C^2 = 0,756$$

producir un efecto pleno ante la presencia de constelaciones de actitudes y orientaciones transmitidas por el sistema educacional. Lo mismo acontecería en el caso de la condición de urbano; estos conceptos se reflejan por los términos del lado derecho del modelo (5). Prosiguiendo con la exploración del problema suscitado, hemos modificado esta línea de argumentación, suponiendo que la condición de urbano es fundamental y que ella mediatiza los impactos provenientes de la economía y del sistema educacional.

Lo que ocurriría sería que la ciudad constituiría una estructura social de naturaleza tal que ella posibilita la operación de ciertas determinaciones económicas y culturales, lo que no acontecería con la estructura social no urbana. Así, por ejemplo, el acceso a bienes de consumo durables puede llegar a determinar un comportamiento reproductivo que se traduzca en una tasa de natalidad menor, pero esta posibilidad, para que sea efectiva, requiere de un mercado de compradores relativamente concentrado. El mismo tipo de argumento es válido en el caso de los servicios de la educación.

Esta línea de argumentación nos conduce a postular el modelo:

$$Y_N = C_0 + C_1 X_1 X_3 + C_2 X_3 X_4 + e_3 \quad (7)$$

Los resultados de su ajuste se presentan en el cuadro 7.

Cuadro 7

RESULTADO DEL AJUSTE DEL MODELO EN QUE LA
URBANIZACION ACTUA INTERACTIVAMENTE CON EL
INGRESO PER CAPITA Y LA EDUCACION BASICA

$$Y_N = 49,844 + 0,000 X_1 X_3 - 0,003 X_3 X_4$$

$$R^2 = 0,781$$

Nuestro intento de incorporar el producto interno bruto per cápita como factor explicativo de la tasa de natalidad, nuevamente ha fracasado.

Algunas voces aisladas ya han señalado que los hallazgos en esta materia no son concluyentes y que la relación entre crecimiento económico y natalidad no constituye una generalización empírica fuera de toda duda. Así, Raúl Urzúa, ^{8/} después de una exhaustiva revisión de la literatura existente indica que:

“Los antecedentes expuestos son suficientes para llegar a algunas conclusiones acerca del apoyo empírico que tiene la creencia de que el crecimiento económico... (conduce) por sí mismo a una disminución de las tasas de natalidad. De atenernos a los resultados obtenidos, debiéramos llegar a la conclusión de que no se ha encontrado una relación clara entre el crecimiento económico... y la tasa de fecundidad predominante en un país”.

Sin embargo, no creemos que los resultados aquí expuestos permitan inferir que no existan determinaciones económicas del comportamiento reproductivo. De hecho, según vimos en la segunda parte de este trabajo, de la estructura de los datos emerge una dimensión que incluye a la variable económica, cuyo efecto sobre la tasa de natalidad es de gran magnitud. Más bien, creemos que nuestros resultados se explican por el carácter mismo de la variable utilizada. Probablemente, la utilización de otros indicadores económicos producirá hallazgos distintos. Así, por ejemplo, es difícil imaginar que variables del tipo de la concentración del ingreso no afecten ese comportamiento. En el fondo, lo que se necesita es información que recupere características estructurales menos gruesas y que permitan detectar los niveles de diferenciación al interior de cada unidad nacional. Basta pensar en el caso de países como Kuwait para percatarse de las deficiencias involucradas por el indicador económico, producto bruto per cápita.

Finalmente, hemos optado por restringir nuestro análisis al modelo:

$$Y_N = D_0 + D_1 X_3 X_4 + e_4 \quad (8)$$

Este modelo se diferencia de los anteriores (excepto el (6), que carece de interacción y que se usa esencialmente para fines de comparación), en que no aparece la variable económica.

El ajuste de (8) ha entregado los resultados que figuran en el cuadro 8.

Es claro que este modelo es el que mejor ajuste presenta, a juzgar por el coeficiente de determinación corregido (R_C^2). A la vez, la magnitud y el sentido del coeficiente asociado a $X_3 X_4$, ha mostrado una constancia notable a través de todo el análisis.

^{8/} Raúl Urzúa, *op. cit.*, pág. 19.

La interpretación sustantiva de este modelo ha sido esbozada anteriormente: la acción sobre el comportamiento reproductivo de las constelaciones de orientaciones y actitudes transmitidas por el sistema educacional ejercen un efecto diferencial, dependiendo del nivel de urbanización alcanzado. Probablemente, es en la ciudad donde las agencias de control social más modernas, tal como el sistema educacional, logran imponerse con mayor fuerza. En síntesis, la ciudad vendría a constituir la condición para que el sistema educacional logre la plenitud de sus efectos.

Cuadro 8

RESULTADO DEL AJUSTE DEL MODELO QUE SOLO
INCORPORA INTERACCION ENTRE EDUCACION BASICA
Y URBANIZACION

$$Y_N = 50,656 - 0,004 X_3 X_4$$

$$(1,829) \quad (0,0005)$$

$$R^2 = 0,773 \quad R_C^2 = 0,760$$

Recapitulando, concluimos que el intento de expresar de manera sintética las relaciones entre las variables explicativas nos condujo a afirmar la existencia de dos factores o dimensiones. Los problemas de interpretación que presenta el segundo factor, en relación con su impacto sobre la tasa de natalidad, nos llevaron a indagar respecto a relaciones de un carácter más complejo: a la luz de los resultados obtenidos en esta parte podríamos aventurar la hipótesis que el efecto de la heterogeneidad socio-económica sobre la tasa de natalidad, simplemente resume diversas interacciones que caracterizan las relaciones de las variables explicativas con el comportamiento reproductivo.

Es posible que estas interacciones se den también respecto a las otras variables no consideradas al interior de los modelos de regresión examinados. En todo caso, hay que destacar que los resultados de los análisis de regresión realizados confirman la conjetura que hicimos al tratar de interpretar el efecto del segundo factor.

Conviene puntualizar de modo somero las principales conclusiones del análisis.

En primer lugar, creemos haber demostrado que el llamado análisis de datos encierra la capacidad de iluminar aspectos complejos de la realidad. Frente a un conjunto de datos que mostraban relaciones recíprocas entre variables, logramos expresar sintéticamente ese conjunto de relaciones, sin reducir de una manera esencial la complejidad. Prueba de ello es que constatamos la existencia de dos

componentes principales capaces de dar cuenta de un altísimo porcentaje de la variación total de la tasa de natalidad.

A su vez, el análisis de componentes principales no significó un cierre prematuro de la indagación, empobreciéndola, sino que, muy por el contrario, las dificultades encontradas al intentar interpretar el efecto del segundo factor nos han abierto toda una línea de reflexión, rica en posibilidades analíticas.

En segundo término, nos parece que los resultados obtenidos indican que es el modelo de los factores el que presenta mayores bondades, entendidas éstas en un doble sentido. Por el lado técnico, es el que muestra un mejor ajuste, a la vez que se acerca más a aquellos *desiderata* implicados por el modelo de regresión.^{9/} Desde otro punto de vista, el modelo de los factores nos permite abandonar el nivel de las variables aisladas para pasar a otro, caracterizado por un lenguaje que utiliza nociones más complejas.

Sin embargo, el análisis de regresión debe entenderse como complementario y como enriquecedor del modelo de los factores: sin él habría sido difícil obtener una interpretación satisfactoria del efecto del segundo factor.

Estas conclusiones deben ser relativizadas a causa de que, por un lado, existen limitaciones propias del análisis de componentes principales y, por otro, hay problemas relativos al nivel de análisis en que nos hemos manejado y la naturaleza de las interferencias aceptables.

Brevemente, destacamos algunas de las limitaciones más importantes del tipo de análisis dimensional utilizado: 1) aun cuando en la realidad no haya relaciones complejas entre variables, la técnica siempre entrega resultados; luego, se requiere como punto de partida alguna evidencia que apunte hacia la existencia de relaciones de esa clase; 2) el análisis de componentes principales es sumamente sensible a las unidades en que se expresen las variables, lo que significa que, por ejemplo, si disponemos de un conjunto de variables medidas en dólares y en libras esterlinas seguramente se obtendrán resultados diferentes, según se use una u otra unidad monetaria; 3) no existen criterios que permitan de una manera automática y enteramente racional decidir acerca del número de factores presentes, y la decisión en este punto debe descansar necesariamente en elementos ajenos a la técnica; 4) por último, hay que destacar que el análisis empleado impone una cierta estructura espacial a los datos y, claramente, es difícil concebir la existencia de razones teóricas u ontológicas que hagan preferir un espacio cartesiano a otras formas posibles de representación.

^{9/} El hecho que los factores no se relacionan linealmente evita problemas de multicolinealidad y, a la vez, puede constituir una salida a su solución. Para un mayor desarrollo de esta idea se puede consultar a J. Jhonston, *op. cit.*, págs. 328-331.

Existen dificultades quizás más serias que las anteriores. Una de ellas se resume usualmente en la literatura mediante la etiqueta de falacia ecológica. ^{10/} El problema reside en la naturaleza de las inferencias que se realizan cuando el análisis utiliza sólo datos que se construyen a partir de unidades agregadas. Se incurre en la falacia en aquellos casos en que se intenta aplicar los resultados del análisis sobre unidades de mayor nivel de desagregación que se encuentran ubicadas en el interior de los conglomerados sobre los cuales se ha obtenido la información. En nuestro caso, caeríamos en la falacia si pretendiéramos reescribir los resultados en términos de comportamiento individual.

Sin embargo, hay que destacar que se abusa bastante de esta noción. Por una parte, no es cierto que toda inferencia respecto de unidades más elementales, derivada de datos agregados, sea ilegítima: en ciertos casos y cumpliéndose ciertas condiciones esas inferencias son válidas. Por otra parte, pensamos que es imposible otorgar algún sentido a generalizaciones empíricas que están referidas a unidades como estados-naciones u otras análogas, sin recurrir a conjeturas sobre la conducta de los individuos. Así, por ejemplo, la relación entre natalidad, población urbana y nivel educacional básico sería bien difícil de interpretar si nos priváramos de utilizar imágenes acerca del comportamiento reproductivo de los individuos. Lo que nos indica la falacia ecológica es que estas conclusiones tienen una naturaleza conjetural y necesitan ser validadas por otros medios.

Finalmente, nos encontramos con la dificultad más grave, que se ubica en la base de este tipo de análisis, vale decir, los problemas que plantea el análisis de cortes transversales. En repetidas ocasiones hemos utilizado las nociones de impacto, efecto y cambio. Comúnmente ellas se piensan en contextos realmente dinámicos: cuando se dice que el nivel educacional básico afecta negativamente la tasa de natalidad, la imagen subyacente se refiere a la trayectoria histórica de una unidad nacional, es decir, se piensa en procesos reales que ocurren en el tiempo.

En cambio, nosotros nos hemos movido en un contexto, por así decirlo, pseudo-dinámico. Nótese que las derivadas parciales, instrumento que permite otorgar sentido a los parámetros de los modelos de regresión examinados, se refieren a variaciones marginales en la tasa de natalidad ante cambios marginales en una de las variables explicativas, pero que éstas nunca han sido definidas respecto a variaciones marginales en la variable tiempo. En otras palabras, cuando hablamos de cambio, en términos estrictos nos estamos simplemente refiriendo a diferencias *simultáneas* entre unidades nacionales.

Desafortunadamente, nuestros intereses y nuestra forma de concebir

^{10/} Para una sistematización del problema y una revisión bibliográfica se puede consultar a: Cortés, Fernando, *El Sistema de Estadísticas Demográficas y Sociales. Aspectos Técnicos. Un Análisis Crítico*. PROELCE, Santiago de Chile, octubre de 1974, págs. 110-123.

la realidad nos lleva, querámoslo o no, a otorgar, aun subconscientemente, un sentido dinámico estricto a las relaciones inferidas de la manera esbozada. Vale la pena, entonces, preguntarse acerca de aquellas condiciones en las cuales esta atribución de contenido dinámico es legítima.

El supuesto que permitiría convertir lo pseudo-dinámico en dinámico implica aceptar que los parámetros de los modelos utilizados no varían a través del tiempo. Vale decir, si realizamos el mismo análisis para un corte transversal de las mismas unidades y variables cinco años más tarde, deberíamos obtener los mismos resultados. Ahora, si pretendemos extrapolar estos resultados a trayectorias nacionales específicas, la dificultad es aún mayor, puesto que debemos suponer que existen leyes de agregación que regulan la inserción de cada unidad en el sistema de diferencias que caracteriza las relaciones entre países.

Creemos que, en general, estos supuestos despertarían fuertes resistencias y prevenciones. El único camino de salida lo constituye el recurrir al estudio de trayectorias históricas específicas, esta vez en un contexto realmente dinámico.

ANEXO 1

MATRIZ DE DATOS *a/*

Países	Producto interno bruto por habitan- te (en US\$ a precios de 1960). 1973		Producto agrícola bruto como porcen- taje del producto interno bruto. 1973		Porcentaje de la po- blación urbana <i>b/</i> 1970	
	X_1	X_1'	X_2	X_2'	X_3	X_3'
Argentina	1 141	2,47	12,3	-0,97	66,4	1,84
Venezuela	935	1,60	6,2	-1,59	56,7	1,28
México	744	0,97	10,6	-1,14	35,3	0,04
Panamá	717	0,87	17,9	-0,41	37,5	0,17
Uruguay	681	0,74	22,1	-0,02	70,2	2,06
Chile	610	0,47	8,1	-1,39	54,6	1,16
Costa Rica	569	0,31	22,0	-0,01	32,2	-0,14
Brasil	447	-0,15	16,9	-0,51	38,5	0,23
Perú	418	-0,26	15,8	-0,62	33,4	-0,07
Colombia	407	-0,30	27,8	0,59	43,1	0,49
Ecuador	388	-0,37	23,0	0,11	32,9	-0,10
Guatemala	383	-0,39	28,1	0,62	17,7	-0,98
Nicaragua	344	-0,54	26,4	0,45	25,3	-0,54
El Salvador	324	-0,61	25,5	0,36	18,1	-0,96
Paraguay	315	-0,65	33,9	1,21	22,5	-0,70
Rep. Dominicana	299	-0,71	18,7	-0,33	27,7	-0,40
Honduras	216	-1,02	34,0	1,22	15,6	-1,10
Bolivia	210	-1,04	20,4	-0,15	22,7	-0,69
Haití	87	-1,51	46,9	2,52	7,0	-1,60

(continúa)

MATRIZ DE DATOS^{a/} (continuación)

Países	Matrícula primaria, como porcentaje de la población de 7-13 años		Matrícula secundaria, como porcentaje de la población de 14-19 años		Alfabetizados, 15 años y más		Tasa de natalidad (por mil habitantes)
	X ₄	X' ₄	X ₅	X' ₅	X ₆	X' ₆	Y _N
Argentina	95,0 ^{c/}	0,66	40,5	0,83	91,4	1,26	21,80
Venezuela	83,4	0,04	38,6	0,70	85,0	0,95	36,06
México	94,4	0,63	24,1	-0,28	77,5	0,58	42,00
Panamá	93,5	0,58	43,0	0,99	78,3	0,62	36,08
Uruguay	95,8	0,71	60,8	2,20	89,4	1,16	20,83
Chile	102,0 ^{d/}	1,04	49,4 ^{e/}	1,43	89,6	1,17	25,88
Costa Rica	101,6	1,02	37,8	0,65	85,8	0,99	33,44
Brasil	76,2	-0,34	35,6	0,50	69,6	0,20	37,12
Perú	104,1	1,15	39,8	0,78	67,0	0,07	41,02
Colombia	62,7	-1,07	20,5	-0,52	72,9	0,36	40,61
Ecuador	86,7	0,22	26,5	-0,12	72,0	0,32	41,76
Guatemala	49,9	-1,75	10,9	-1,17	37,9	-1,43	42,82
Nicaragua	74,0	-0,46	19,5	-0,59	49,8	-0,76	48,32
El Salvador	76,1	-0,35	19,6	-0,58	50,8	-0,72	42,17
Paraguay	92,6	0,54	17,6	-0,72	69,0	0,17	42,25
Rep. Dominicana	89,3	0,36	19,4	-0,60	53,1	-0,60	45,84
Honduras	76,6 ^{e/}	-0,32	12,7	-1,05	47,0	-0,90	49,26
Bolivia	84,6	0,11	16,0	-0,83	39,8	-1,25	43,72
Haití	31,0 ^{e/}	-2,76	4,2 ^{e/}	-1,62	18,8	-2,27	41,98

Fuentes: Para las variables demográficas se ha usado Naciones Unidas, CEPAL, Situación Demográfica Alrededor de 1973 y Perspectivas para el Año 2000. ST/ECLA/CONF. 48/L.5, 25 de marzo de 1971, cuadro 1, pág. 3. Para las variables económicas y sociales, Naciones Unidas, CEPAL, *Economics Survey of Latin America, 1973.*

- a/ X_i se refiere a las variables originales y X'_i a las mismas, pero una vez tipificadas.
- b/ Población urbana definida como aquella que vive en ciudades de más de 20 000 habitantes.
- c/ 6-13 años de edad.
- d/ 7-15 años de edad.
- e/ 1968.