Análisis Exploratorio Multivariado. La conformación de subregiones al interior de cuatro provincias argentinas según el impacto del desarrollo en las condiciones de vida

Silvia Lago Martínez. Socióloga. Profesora Adjunta de Metodología de la Investigación Social. Carrera de Sociología. Facultad de Ciencias Sociales. Universidad de Buenos Aires e Investigadora del Instituto Gino Germani de la misma Facultad.

Mirta Mauro. Socióloga. Docente de la misma Cátedra y de Estadística de la Facultad de Psicología de la UBA.

Gustavo Alvarez. Sociólogo. Docente de la misma Cátedra y Analista de Instituto Nacional de Estadísticas y Censos (INDEC)

Abstract

This study conforms an essay, in the terms of an investigation of a faculty subject at University, on the usage of exploratory methods of multivaried analysis. The exercise consists in conforming a regionalisation of subprovincial areas according to their level of social development. In order to achieve this work is being done with data on the argentine provinces of Mendoza and San Juan (Cuyo region) and La Rioja and Catamarca (the argentine North West). The geographical units taken into account are departments, considered these as political and administrative divisions. The focus is set on pointing out that, in spite of the existent disequilibriums in the levels of economic development of the Cuyo regions (which show a greater degree of development) and the argentine North West (showing a lesser degree of development), to the inside of each regions or province situations are not homogeneous, and it is possible to detect departments in different provinces whose life conditions -in relation to a group of indicators- are similar or likely to be grouped in differentiated zones. By combining multivaried methods (factorial analysis and cluster) complex variables are reduced and units (departments) are classified into "strata", starting from the construction of a complex variable "Impact of Economic Development on Life Conditions" (IEDLC), according to a group of demographic, social and economic indicators, which have already been published. These resulting strata, in fact, have recognized affinities among departments of different provinces. That is to said, that they show a certain composition in their social development which crosses over provincial limits and difference among regions.

Key words: multivaried analysis, regionalisation of subprovincial areas, level of social development.

Introducción

Este trabajo propone la utilización de métodos exploratorios multivariados, que permitan producir información orientativa para la reducción de variables complejas y la clasificación de unidades en el marco de esa complejidad.

El propósito es arribar a una clasificación de unidades geográficas -divisiones político administrativas- de las Provincias de Mendoza, San Juan, Catamarca y La Rioja, que de cuenta del Impacto del Desarrollo en las Condiciones de Vida de la población de estos Departamentos y permita conformar una regionalización de áreas subprovinciales.

A tal efecto, se utilizaron veintiséis indicadores referidos a diferentes dominios temáticos -sociales, económicos y demográficos- comparables y ya publicados. Las fuentes utilizadas son las publicaciones del INDEC "Situación y Evolución Social Provincial" de cada una de las Provincias, que aglutina datos del Censo Nacional de Población y Vivienda 1991, series B, C y tabulados inéditos, de las Secretarías de Estado de Salud Pública de las Provincias y del Ministerio de Salud y Acción Social de la Nación.

La metodología utilizada consiste en la aplicación de un análisis tendiente a reducir el espacio multidimensional propio de los datos originales, combinando dos técnicas de análisis espacial. Se compone de dos pasos:

1. En primera instancia se realiza un análisis factorial para reducir un número elevado de indicadores (25) a una cantidad discreta de dimensiones que los refleje y al mismo tiempo, resuma la información, habida cuenta de la elevada interrelación entre muchos de aquellos indicadores. Se identifican seis componentes principales (dimensiones) con una capacidad explicativa relativamente importante.

2. Posteriormente se clasifican los departamentos de las cuatro provincias analizadas -setenta y uno en total- en una cantidad reducida de estratos con cierta homogeneidad interna y elevada diferenciación entre ellos.

A tal fin, se utiliza una técnica de conglomerados, por medio del procedimiento de k-medias, teniendo como insumo el valor de las cinco principales dimensiones extraídas durante el análisis factorial (el indicador que se tomó fue el puntaje de cada factor), se arriba a una clasificación de los departamentos en cuatro estratos, con diverso Impacto del Desarrollo en las Condiciones de Vida.

Finalmente se evalúa el resultado de esta estratificación midiendo la asociación con el porcentaje de hogares con NBI. En tal sentido, se realizaron análisis de varianza para corroborar que, si bien la provincia de pertenencia de los departamentos está asociada con la incidencia del NBI, las variaciones de este indicador son más explicados por el estrato definido en este ejercicio.

1. El Estado de la Cuestión y la Caracterización de las Provincias

Este apartado tiene por objeto caracterizar a las Provincias analizadas a partir de algunas variables macro-económicas, que entre otras, explican el nivel de "desarrollo" de una región o Provincia (Rofman 1995, Mayo 1995, Manzanal 1995), entendiendo desarrollo regional como "desarrollo económico y social de divisiones territoriales o grupos ecológicos de una sociedad nacional" (P. Heintz 1970). El interés está dado en señalar, precisamente, las diferencias en los niveles de desarrollo económico entre las regiones de Cuyo -especialmente Mendoza y San Juan- y Noroeste -Catamarca y La Rioja- que impactan en las condiciones de vida de sus poblaciones.

Diversos especialistas en el estudio de las economías regionales toman en cuenta la existencia de dos ámbitos claramente diferenciados del territorio nacional: la región central o pampeana y el conjunto de provincias extra-pampeanas. Este carácter diferenciado y jerárquico de la estructura económica se encuentra estrechamente vinculado con el proceso de su formación histórica y la matriz de los modelos de desarrollo nacional, al mismo tiempo que las transformaciones económicas producidas a partir de 1990 en lugar de desdibujar las diferencias, acentuaron el desequilibrio regional, perjudicando especialmente a las Provincias cuyo eje de acumulación está focalizado en el mercado interno (A. Mayo 1995).

Según una clasificación de la Secretaría de Programación Económica (1994) que toma como base el promedio anual de exportaciones de 1993-94 surgen cuatro grandes grupos de Provincias:

"grandes" (ventas anuales superiores a los 1.000 millones de dólares) Bs. As./ Ca. Federal, Santa Fé y Córdoba.

"medias altas" (de 250 a 990 millones), Chubut, Mendoza, Neuquén y Santa Cruz.

"medias-bajas" (de 50 a 249 millones), Entre Ríos, Tierra del Fuego, Río Negro, Tucumán, Salta, Misiones, Chaco, La Pampa, San Luis, Corrientes, Jujuy, La Rioja y Santiago del Estero.

"bajas" (inferiores a 50 millones), Formosa, San Juan y Catamarca.

Dentro de la región Cuyo, la Provincia de Mendoza y la de San Juan, presentan situaciones muy diferenciadas. Por un lado Mendoza muestra una considerable diversificación y San Juan un claro estancamiento (Cuadro 1). Siguiendo la caracterización de Mayo (1995), Mendoza desarrolló históricamente una importante producción vitivinícola destinada al mercado interno que permitió crear una cierta base manufacturera y absorber la fuerza de trabajo de la región. Además de la vid, Mendoza es uno de los principales centros de producción y de transformación frutícola y hortícola del país y tiene una destacada inserción en mercados externos, como el Brasil. En 1994 presentaba una tasa relativamente baja de desempleo (5,1%).

San Juan, por el contrario, presenta una situación muy distinta a la de Mendoza y aún a la de San Luis (sus pares de Cuyo). Las exportaciones permanecen estancadas y ello puede ser atribuido, tanto a la crisis del sector vitivinícola, como a su bajo desarrollo agrícola y también al fracaso de obtener una expansión manufacturera a través del régimen de promoción industrial.

El NOA es la región argentina que presenta el panorama más grave. En Catamarca y La Rioja se registra una situación de virtual monocultivo (monocultivo de olivo) y con índices muy elevados de propiedad minifundista. Es también, la zona extra-pampeana con los índices más altos de desocupación del país (Cuadro 2). En Catamarca y la Rioja las estructuras industriales son muy débiles y el empleo público figura entre los más elevados del país (el 9,6% en relación con la PEA en La Rioja y el 8,5% en Catamarca, en 1994). Catamarca es la provincia argentina con menor inserción en el mercado mundial y sus exportaciones se reducen a algunos productos agrícolas, muchos de los cuales no son despachados directamente de su territorio. La Rioja, si bien aumentó extraordinariamente sus exportaciones, esta expansión se debe a una muy fuerte concentración en una sola actividad (curtiembre), mientras la mayoría de la producción es de monocultivo y sumamente atrasada.

Estos datos se yuxtaponen a una gran crisis fiscal. A partir de 1995 cuando los presupuestos provinciales caen, las provincias se ven obligadas a endeudarse, y a afectar su coparticipación impositiva. Estudios recientes (Vaca y Cao 1998) caracterizan la generación de la deuda de los estados provinciales como producto de un persistente incremento del gasto, originado en urgencias (pago de sueldos, jubilaciones, etc.) que, en el mediano plazo, devenían en desequilibrios financieros-fiscales más profundos que los que había originado el endeudamiento. En julio de 1998 (informe económico-regional del MeyOySP) el stock de deuda pública de las Provincias superaba los 16 millones de pesos, y el stock de deuda provincial como porcentaje de recursos totales era de 51,9% como promedio. Todas las Provincias en estudio superan este promedio: Mendoza 94,8%, La Rioja 83,1%, Catamarca 72,3% y San Juan 55,1%.

Por último, es interesante comentar un trabajo reciente de Rofman (1997) donde realiza una evaluación del impacto del plan de convertibilidad sobre las 28 aglomeraciones urbanas que releva el INDEC (por medio de la EPH), en el período 1985-1995. Construye una tipología combinando factores de comportamiento de las estructuras de producción y de empleo (1) y obtiene los siguientes "tipos":

1. Aglomeraciones especializadas en actividades vinculadas al aparato burocrático-administrativo del Estado.

En este tipo sobresale la presencia del aparato del estado, complementado con otras actividades del sector terciario.

2. Aglomeraciones caracterizadas por una combinación de actividad burocrática oficial importante con nuevos emprendimientos industriales o de servicios.

Aquí incluye a La Rioja, Catamarca y San Juan, señalando que las políticas de promoción no se inscribieron en un contexto mayor sobre las bases de un plan nacional de desarrollo regional, sino que surgieron como iniciativas puntuales que no presentan signos de estabilidad.

3. Aglomeraciones que se destacaron como centros industriales significativos o poseyeron actividades extractivas dinámicas en períodos previos y hoy enfrentan una seria declinación.

4. Aglomeraciones que lograron captar nuevas actividades industriales o de intermediación en las últimas décadas sin contar con beneficios estatales.

Aquí señala al Gran Mendoza como muy representativa (al igual que el Gran Córdoba) de esta situación, con procesos de transformación y de intercambio Comercial importante, y un mercado laboral más estable.

5. Aglomeración mayor del país relevada como una sola unidad aunque posee fuertes diferencias económicas y sociales a su interior: el área metropolitana del Gran Buenos Aires.

2. El Impacto del Desarrollo Económico en las Condiciones de Vida

Tal como señalamos arriba, las regiones -y aún las Provincias dentro de las regiones- en que se encuentra dividido el espacio nacional, presentan diversos grados de desarrollo. No es el propósito de este trabajo introducir el debate acerca de los "desequilibrios" o "desigualdades" regionales (A. Rofman 1974), que apuntan a señalar la heterogeneidad del desarrollo socio-económico de las mismas. Por el contrario, el objetivo es clasificar las unidades político administrativas (departamentos) de las Provincias en estudio, en subregiones, con cierta homogeneidad interna y elevada diferenciación entre ellos, que atraviesen los límites provinciales.

Sin duda, los signos de, retraso, estancamiento o desarrollo productivo propio, que caracterizan a cada Provincia impactan sobre las condiciones de vida de la población. Sin embargo, al interior de cada una de ellas, éstas situaciones no son homogéneas y es posible detectar departamentos de diferentes provincias cuyas condiciones de vida -con relación a un conjunto de indicadores- sean similares o plausibles de ser agrupadas en zonas diferenciadas.

Estas subregiones se clasifican como "estratos" a partir de la construcción de una variable compleja que se denomina "impacto del desarrollo económico en las condiciones de vida (IDECV)", de acuerdo a un conjunto de indicadores demográficos, sociales y económicos, que hacen referencia a factores de orden estructural propios del sistema económico social vigente en la Argentina.

Las áreas temáticas - y sus respectivos indicadores- consideradas son (2):

a. Status urbano rural (indicador: porcentaje de población urbana)

b. Condiciones migratorias (indicadores: porcentajes de migrantes internos interprovinciales y porcentaje de migrantes de países limítrofes).

c. Conformación de los hogares (indicadores: promedio de personas por hogar y porcentaje de jefa de hogar mujer)

d. Pobreza (indicador: porcentaje de hogares con NBI)

e. Precariedad de la vivienda y de los servicios asociados a ella y sus condiciones de tenencia (indicadores: porcentaje de viviendas precarias, porcentaje de viviendas sin agua corriente, porcentaje de viviendas sin electricidad, porcentaje de hogares sin gas de red o envasado, porcentaje de población en hogares con servicios de desagüe a red cloacal, porcentaje de hogares en situaciones irregulares de tenencia).

f. Condiciones de salud y cobertura de salud (Indicadores: tasa de mortalidad infantil total, tasa de mortalidad neonatal, porcentaje de madres de hijos nacidos vivos menores de 20 años y porcentaje de población sin cobertura en salud)

g. Escolarización y analfabetismo (Indicadores: tasa neta de escolarización de nivel medio, porcentaje de población que ya no asiste pero asistió de nivel hasta primario incompleto y tasa de analfabetismo)

h. Estructura del empleo, desocupación y precariedad laboral (Indicadores: tasa de actividad económica total, tasa de actividad económica de mujeres, tasa de desocupación total, tasa de desocupación de mujeres, porcentaje de población ocupada en el sector público, porcentaje de trabajadores por cuenta propia y porcentaje de asalariados sin descuentos previsionales).

Los datos utilizados en cada caso corresponden a relevamientos que surgen de diferentes fuentes para el año 1991. Se toma este año, con el objeto de homogeneizar temporalmente la información, dado que para la mayoría de los indicadores a nivel interprovincial, no se registran datos más actualizados. Si bien para la caracterización de las regiones en cuestión, se cuenta con información más reciente, la misma hace referencia a procesos de orden estructural y a sus formas de manifestación en las áreas bajo estudio. Los fuertes cambios económicos y políticos a partir de 1991 no han hecho más que ahondar los desniveles de desarrollo entre las regiones que ya se encontraban planteados a principio de la década.

Cuadro 1. Cuyo, Exportaciones y Tasa de Desempleo (en millones de dólares y en porcentajes)
Exportaciones en millones de dólares

	1990	1991	1992	1993	1994
Mendoza	426	371	351	384	452
San Juan	39	33	43	29	38
San Luis	41	43	64	96	122

Desempleo en porcentajes

	1990	1991	1992	1993	1994
Mendoza	5,8	4,4	4,4	4,6	5,1
San Juan	9,7	7,0	9,3	6,0	8,1
San Luis	4,7	5,3	6,7	8,9	6,7

Fuente: Mayo A. (1995) información de la Secretaría de Promoción Económica, 1994.

Cuadro 2. Noa, Exportaciones y Tasa de Desempleo (en millones de dólares y en porcentajes)
Exportaciones en millones de dólares

	1990	1991	1992	1993	1994
Catamarca	6	10	12	10	11
Jujuy	149	105	91	78	74
La Rioja	16	21	34	48	63
Salta	229	211	172	182	180
Stgo. del Estero	44	53	42	36	67
Tucumán	181	178	166	153	229

Desempleo en Porcentajes

	1990	1991	1992	1993	1994
Catamarca	9,2	9,0	5,9	7,9	7,5
Jujuy	5,7	3,3	6,6	6,5	9,7
La Rioja	6,1	5,7	8,7	5,7	10,9
Salta	7,7	4,7	9,8	10,1	13,3
Stgo. del Estero	2,8	3,2	2,3	3,9	4,6
Tucumán	9,4	11,4	12,5	11,8	14,2

Fuente: Mayo A. (1995) Secretaría de Promoción Económica, 1994.

3. Estratificación de Departamentos según el IDECV

Este ejercicio se propuso explorar la posibilidad de identificar espacios subregionales, esto es configuraciones socioespaciales que atraviesen estas provincias. En tal sentido, se aspiró a reconocer estratos espaciales que sean más reveladores de condiciones equivalentes. Por ello, se apuntó a definir agrupamientos de unidades espaciales intraprovinciales, departamentos, que fuesen significativos. La meta fue que dichos agrupamientos cumpliesen dos condiciones: ser consistentemente homogéneos en su interior y considerablemente distintos a los restantes estratos.

Dado que las unidades de análisis correspondían a cuatro contextos, es decir cuatro provincias, se exploraron diversas formas de estratificar a los departamentos en esa misma cantidad de grupos. El supuesto en el que se basó esta decisión fue que sería la forma más adecuada de contrastar la hipótesis nula. En efecto, el análisis exploratorio propende a identificar el agrupamiento óptimo, esto es aquel que brinda la mejor predicción de la variabilidad de la información. Ahora bien, si ese agrupamiento hubiera coincidido con el trazado provincial deberíamos haber asumido que las diferencias interprovinciales son de tal entidad que impiden hallar espacios subregionales.

A los efectos de contar con una base amplia de comparaciones entre los departamentos de las provincias observadas, se analizó un conjunto de veinticinco indicadores. Sin embargo la utilización de esos valores no se dio sin un examen previo. Al respecto, se aplicó un análisis factorial con la finalidad de orientar la exploración temática en dos sentidos:

- hallar dimensiones implícitas en la variedad de las mediciones;

- reducir la complejidad del espacio de propiedades a sus componentes principales.

Aplicar esta modalidad de análisis de la información disponible conlleva importantes ventajas. En general, los autores que describen al análisis factorial señalan que:

- es una representación más parsimoniosa, ya que permite representar numerosos datos por medio de un número más reducido de parámetros numéricos;

- los parámetros derivados de la reducción, factores, se basan en un conjunto amplio de datos y por tal motivo exhiben una mayor fiabilidad estadística (3).

El resultado del análisis factorial fue el reconocimiento de cinco dimensiones, expresadas en los respectivos puntajes factoriales. Sobre la base de estos cinco índices sintéticos se procedió a explorar diversas estratificaciones de los departamentos valiéndose del método de análisis de conglomerados (o de clusters).

El propósito del análisis de conglomerados es agrupar a un conjunto de unidades, teniendo en cuenta las similitudes que las mismas presentan en un número elevado de variables. Los conglomerados serán tales que la

distribución de los valores de las variables en cada uno sean lo más homogénea posible y, al mismo tiempo, los conglomerados sean muy distintos entre sí.

El análisis de conglomerados se basa en el cálculo de distancias medias de las unidades de análisis a través de un número considerable de mediciones. La forma más simple de establecer la semejanza de dos unidades es a través de la distancia euclídea entre ellas. La misma se obtiene por la raíz cuadrada de la sumatoria de las diferencias cuadráticas entre las unidades en cada dimensión (4):

d_{ij} = Ö $(X_i-X_j)^2+(Y_i-Y_j)^2+...+(Z_i-Z_j)^2$

Siendo:

i: primera unidad de análisis

j: segunda unidad de análisis

X, Y, Z: dimensiones o variables

Sobre la base de estas diferencias se confecciona una matriz de distancias entre las unidades. En el caso particular de este ejercicio, el cálculo de las distancias medias entre las unidades (los departamentos) se llevó a cabo con los puntajes factoriales. Esta modalidad comporta dos ventajas:

- dado que se calculan distancias medias, el uso de las cinco dimensiones en lugar de los veinticinco indicadores supone un tratamiento más puro de las diferencias ya que se eliminan las repeticiones que se derivan del uso de dos o más indicadores de una misma dimensión y se pondera a cada una en la misma medida;

- siendo que las distancias se establecen en unidades de cuenta de la respectiva variable, comparar puntajes factoriales brinda una base de igualación ya que estos puntajes corresponden a distribuciones estandarizadas impidiendo que las diferencias sean sensibles a las variables con escalas de puntuaciones mayores.

Finalmente el procedimiento de aglomeración concretó una estratificación de los departamentos en cuatro categorías. A los efectos de evaluar la utilidad de esta clasificación se comprobó que la misma no coincidía con el agrupamiento por provincias. Por otra parte, se validó la estratificación con una variable externa confirmando que el resultado de la aglomeración era más satisfactorio que la identificación de los departamentos por su provincia de pertenencia.

3.1 Identificación de Dimensiones del Impacto del Desarrollo en las Condiciones de Vida (IDECV): Análisis Factorial de Componentes Principales

El punto de partida del análisis fue una matriz de 25 indicadores (5) para 71 casos. Los indicadores utilizados pertenecían a distintas áreas temáticas, refiriéndose a cuestiones tales como la situación habitacional, la participación en el mercado laboral, el acceso al sistema educativo formal y las condiciones de salud. Los mismos fueron reunidos en razón de que podrían dar cuenta del impacto del desarrollo en las condiciones de vida (6).

En tal sentido, el interrogante a resolver era ¿cuáles eran las afinidades e interrelaciones que existían entre esos indicadores? Responder a este cuestionamiento fue la base para simplificar la representación de los datos, ya que un conjunto limitado de dimensiones principales podía reproducir la información original soslayando aspectos redundantes.

La medida de las interrelaciones entre los indicadores analizados se obtuvo por medio del coeficiente momento-producto de Pearson, tradicionalmente conocido como r. Dicha matriz fue acompañada por la correspondiente a la medición del nivel de significación de aquellas correlaciones (7). Estos resultados podrían haber sido analizados

minuciosamente, mas en razón de su magnitud y complejidad se apeló a un método de extracción factorial para identificar grupos de indicadores afines.

A los efectos de reducir los datos, simplificando los veinticinco indicadores originales en un conjunto menor de variables, se utilizó el método de extracción de componentes principales. El mismo transforma una serie de variables en un conjunto de nuevas variables denominadas componentes principales. Los componentes principales resultan de la combinación lineal de las variables originales maximizando la varianza explicada de los datos iniciales. En definitiva, el primer factor (o primer componente principal) es una nueva variable generada por la regresión lineal múltiple a partir de todas las variables existentes, bajo el requisito de que la variación explicada sea la máxima posible (con un valor de r^2 que sea superior a cualquier otro hallable por medio de otras ecuaciones con los datos analizados) (8).

$F_1 = b_1 X + b_2 Y + ... + b_m Z$

Siendo:

F_1 : Primer factor

b_1, b_2, b_m: coeficientes de regresión de cada variable para el primer factor

X, Y, Z: variables

Posteriormente se resta la varianza explicada por este factor y se extrae con el mismo procedimiento un nuevo componente que tendrá la característica de ser el que mejor explica la varianza en la matriz residual. El procedimiento se desarrolla de manera que cada nuevo componente completa la explicación de la matriz de datos hasta dar cuenta de la variabilidad total de los datos.

Cada nuevo componente, cumple con el requisito de no estar correlacionado con el factor anterior. Por tal motivo, esta nueva combinación lineal es perpendicular a la obtenida en el paso precedente.

El objetivo de explicar la variabilidad total recién se logra cuando se reconocen tantas combinaciones lineales como variables había en la matriz inicial. Ahora bien, como el método entrega resultados jerarquizados se aprecia que los primeros componentes concentran una proporción importante de la variabilidad de los datos. Frecuentemente, y esa es la utilidad que se deriva de la extracción de componentes principales, una cantidad reducida de factores (combinaciones lineales de las variables) explica una proporción elevada de la variabilidad.

La extracción de componentes principales, a partir de los veinticinco indicadores originales, derivó en la selección de seis factores (9). Entre ellos, daban cuenta del 82,5% de la variabilidad total de los datos (10). Es decir que se podría expresar aproximadamente la misma variedad, entre las unidades analizadas, con seis factores principales que con veinticinco indicadores.

Este resultado da cuenta de la existencia de una elevada intercorrelación entre las variables, esto es que en muchos casos varios indicadores remiten a la misma dimensión. A fin de caracterizar a las dimensiones implícitas se analizó la correlación de cada indicador con los factores extraídos. La meta era distinguir claramente cuáles indicadores se asociaban con cada factor. En tal sentido, se procedió a rotar los factores (11) en un procedimiento que permite intensificar la diversidad entre los distintos componentes sin alterar la configuración espacial de los datos iniciales.

Dado que los indicadores correlacionados tendrían un comportamiento común, se intentó establecer una noción representativa del significado conjunto que tendrían los indicadores. En el esquema 1, se detalla la composición de cada factor (12) aclarando con un signo menos entre paréntesis los indicadores que tienen una correlación de sentido negativo con el factor común.

Esquema 1. Impacto del Desarrollo en las Condiciones de Vida. Composición Factorial

Factor 1 Acceso deficitario a la educación y la vivienda

% Viviendas sin agua corriente
Tasa Neta de Escolarización del Nivel Medio **(-)**
%Viviendas Precarias
% de Viviendas sin Electricidad
% de Adultos que sólo asistieron hasta el nivel primario incompleto
% de Población Urbana **(-)**
Tasa de Analfabetismo
% de Hogares sin gas de red o envasado

Factor 2 Acceso al mercado laboral

Tasa de Desocupación de Mujeres **(-)**
Tasa de Desocupación de Ambos sexos **(-)**
Tasa de Actividad de Ambos sexos
Tasa de Actividad de Mujeres
% de Hogares con Tenencia Irregular de Vivienda **(-)**
% de Migrantes Internos Interprovinciales

Factor 3 Inestabilidad sociolaboral

% de Asalariados Precarios (sin descuento previsional)
% de Asalariados del Sector Público **(-)**
% de Migrantes de Países Limítrofes
% de Hogares con Jefatura Femenina **(-)**

Factor 4 Mortalidad infantil

Tasa de Mortalidad Neonatal
Tasa de Mortalidad Infantil

Factor 5 Condiciones riesgosas para la infancia

% de Nacidos con madres de menos de 20 años de edad
% Viviendas con desagüe cloacal a red pública **(-)**
Promedio de personas por hogar

Factor 6 Inserción incompleta al mercado laboral

% trabajadores por cuenta propia
% de población sin cobertura de salud

3.2 Estratificación de los Departamentos de Cuatro Provincias según el IDECV: Análisis de Conglomerados por Método de *k*-medias

El agrupamiento de los departamentos analizados se concibió para que contuviese el carácter complejo del IDECV, al tiempo que resultase en cuatro estratos significativos. La clasificación debía ser tal que cada estrato tuviese una elevada homogeneidad interna y una alta heterogeneidad en referencia a los demás.

El resultado del análisis factorial indicó la presencia de seis dimensiones constitutivas del IDECV. No obstante ello, un análisis minucioso llevó a descartar el último de los factores, esto es la dimensión definida como 'Inserción incompleta en el mercado laboral'. Entre otros aspectos, se tuvo en cuenta que los indicadores que la componían no estaban intensamente correlacionados y que el valor de variabilidad aportado por este elemento, si bien superior, estaba demasiado próximo al punto de corte. Por otra parte, eliminando este factor la comunalidad de los cinco restantes no se veía notoriamente menguada. En efecto, los cinco factores principales daban cuenta del 77.7% de la variabilidad total de los datos (13).

En consecuencia, la aglomeración se llevó a cabo sobre la base de los puntajes factoriales derivados de los cinco componentes principales. Para establecer la similitud entre las unidades se utilizaron las distancias euclídeas. Mientras que para determinar la asignación de las unidades a los grupos se empleó el método de las *K*-medias.

El método de aglomeración de las *K*-medias aglutina las unidades en un número *k* de grupos que es definido *a priori*. Para ello se define una cantidad *k* de puntos centrales o centroides; el grupo de pertenencia de cada unidad se define por el centro con el cuál se halla la menor distancia.

La definición de los *k* puntos puede fijarse con un criterio previo o bien puede ser estimado a partir de los datos disponibles. Cuando se pretende estimar los *k* centroides, se utiliza un procedimiento iterativo. En un primer paso se adoptarán los *k* primeros casos de la matriz de datos como centros iniciales temporales. A partir de ellos, se tratará de mejorar la solución siguiendo paso por paso la siguiente regla: si la menor distancia de un caso a un centro es mayor que la menor distancia entre los dos centros más cercanos, se sustituirá el centro más próximo.

Una vez estimados los *k* centroides iniciales, el paso siguiente es calcular la distancia de cada punto a cada uno de ellos. En función de la mínima distancia obtenida, los puntos se agruparán en *k* conglomerados. Cada grupo estará formado por el elemento que dio lugar al centroide inicial y todos aquellos que la distancia a dicho centro sea la mínima entre las correspondientes *k* posibles.

Cuando todos los individuos han sido agrupados en *k* conglomerados, se calculan los centros para cada uno de los grupos. El centroide del conglomerado es el punto definido por las medias de las variables para los elementos que conforman el grupo. A partir de estos centros de conglomerados, se reagrupan los individuos obteniendo un nuevo agrupamiento. Dado que el proceso de iteración puede demorar varias pasadas, a priori se define un criterio de convergencia que puede estar dado por la cantidad máxima de iteraciones o bien por una proporción de cambio en referencia a los centros iniciales (14).

El procedimiento de aglomeración aplicado se basó en cinco puntajes factoriales correspondientes a las principales dimensiones del IDECV. Dichos datos se utilizaron para conformar cuatro estratos de departamentos mediante el método de *K*-medias. Dado que por cuestiones teóricas se pretendía conformar cuatro grupos se trabajó con cuatro centroides que fueron inicialmente estimados a partir de la divergencia propia de los datos. Si bien se había fijado un máximo de diez iteraciones, se arribó a una solución satisfactoria en sólo siete pasos (15).

La estratificación de los departamentos según el IDECV se plasmó en cuatro grupos de desigual tamaño. Como se aprecia en el Cuadro 1, mientras el Estrato III concentra casi la mitad de los casos, el Estrato IV apenas representa el diez por ciento.

Cuadro 1. Distribución de departamentos según IDECV

	Frecuencia	Porcentaje
Estrato I	14	19.7
Estrato II	17	23.9
Estrato III	33	46.5
Estrato IV	7	9.9
Total	**71**	**100.0**

Fuente: Anexo Estadístico Cuadro 7

La caracterización de cada estrato se desarrolló sobre la base de la puntuación media obtenida para cada dimensión. En tal sentido, pueden establecerse las siguientes apreciaciones para caracterizar a cada uno en términos relativos:

Estrato I Es un conjunto de departamentos que tienen un acceso relativamente adecuado a la educación y la vivienda, además poseen una mortalidad infantil levemente superior al promedio regional y valores inferiores al promedio en accesibilidad al mercado laboral pero con una baja inestabilidad sociolaboral;

Estrato II Son departamentos que tienen una elevada inestabilidad sociolaboral, por otra parte, tienen baja incidencia de condiciones riesgosas para la infancia y de mortalidad infantil y una accesibilidad al mercado laboral elevada

Estrato III Los departamentos de este estrato no tienen un perfil tan marcado como los anteriores. No presentan valores muy deficitarios en cuanto acceso a la educación y la vivienda ni en cuanto a la mortalidad infantil, en cambio, ostentan un mercado laboral con baja accesibilidad aunque con baja inestabilidad laboral

Estrato IV Este es el estrato más definido ya que presenta un elevado déficit en el acceso a la educación y la vivienda junto con una alta mortalidad infantil. Por otra parte, exhibe una importante accesibilidad al mercado laboral y una baja inestabilidad sociolaboral (16).

Más allá del perfil particular de los estratos generados por la clasificación, el propósito del ejercicio era arribar a una estratificación que fuese más significativa que la distinción en provincias para caracterizar a los departamentos. En tal sentido, la primera comprobación que se hizo fue que la estratificación de IDECV no coincide totalmente con la provincia de pertenencia de los departamentos.

Cuadro 2. Noa, Exportaciones y Tasa de Desempleo
Exportaciones en millones de dólares

	1990	**1991**	**1992**	**1993**	**1994**
Catamarca	6	10	12	10	11
Jujuy	149	105	91	78	74
La Rioja	16	21	34	48	63
Salta	229	211	172	182	180
Stgo. del Estero	44	53	42	36	67
Tucumán	181	178	166	153	229

Desempleo en porcentajes

	1990	**1991**	**1992**	**1993**	**1994**
Catamarca	9,2	9,0	5,9	7,9	7,5
Jujuy	5,7	3,3	6,6	6,5	9,7
La Rioja	6,1	5,7	8,7	5,7	10,9
Salta	7,7	4,7	9,8	10,1	13,3
Stgo. del Estero	2,8	3,2	2,3	3,9	4,6
Tucumán	9,4	11,4	12,5	11,8	14,2

Fuente: Mayo, A 1995. Secretaría de Promoción Económica 1994.

En el cuadro 2, se advierte que en ningún caso hay una estricta correspondencia entre un estrato y una provincia. En general, se halló que la mayor parte de los departamentos de una provincia se concentraron en un estrato.

Justamente el Estrato III, que es el más numeroso, concentra alrededor de un tercio de los departamentos de Catamarca, La Rioja y San Juan.

En cambio, los estratos suelen tener una composición que atraviesa las fronteras provinciales. Sólo el Estrato II se compone en una elevada proporción por departamentos de una sola provincia. Opuestamente a éste, el Estrato IV se nutre por partes iguales por departamentos de La Rioja y Catamarca, mientras que el Estrato I lo conforman similares cantidades de departamentos de Mendoza, La Rioja y San Juan.

Dado que se comprobó cierta independencia entre la provincia de pertenencia y los estratos de IDECV, restaba verificar que el agrupamiento creado constituía una taxonomía reveladora para el análisis de espacios subregionales. En tal sentido, se estableció un ejercicio de validación externa para comparar las dos clasificaciones disponibles (Estrato de IDECV y provincia de pertenencia).

Para el desarrollo de la validación se utilizó un indicador de incidencia de la pobreza, el % de Hogares con Necesidades Básicas Insatisfechas (NBI) que reunía dos condiciones propicias para este examen:

- era un referente de condiciones de vida que no había formado parte de la medición multivariada del IDECV;

- se trataba de un indicador complejo en sí mismo ya que había sido definido por la combinación de cinco indicadores (17). La técnica empleada para evaluar la asociación de cada agrupamiento con el indicador de pobreza, fue el análisis de varianza. El supuesto que se pretendía comprobar era que la aglomeración de los departamentos según el Estrato de IDECV tenía mayor relación con los '% de Hogares con NBI', que el recorte según provincia de pertenencia.

Cuadro 3. Resultados del análisis de varianza del 'Porciento de hogares con NBI' según Provincia de pertenencia y según Estrato de IDECV

	Asociación Eta	**Determinación Eta-cuadrado**
Provincia de pertenencia	.5302	.2811
Estrato de IDECV	.8189	.6706

Fuente: Anexo Estadístico Cuadro 10

La síntesis del análisis de varianza se aprecia en el Cuadro 3. Según los valores que se hallaron, la clasificación construida por el proceso multivariado resultó más asociada a la incidencia de la pobreza que la distinción de los departamentos por provincia. La estratificación concebida ad-hoc exhibió un importante nivel de asociación (Eta= .8189), al punto que esta clasificación de los departamentos explica alrededor de dos terceras partes de la variación total de la incidencia de la pobreza entre ellos.

4. Comentarios Finales

El análisis factorial de componentes principales cumplió con el objetivo de reducir la elevada cantidad de indicadores iniciales a un conjunto reducido de dimensiones. En tal sentido, entre los veinticinco indicadores analizados se halló seis dimensiones principales conformando el IDECV:

1. Acceso deficitario a la educación y la vivienda
2. Acceso al mercado laboral
3. Inestabilidad sociolaboral
4. Mortalidad infantil
5. Condiciones riesgosas para la infancia
6. Inserción incompleta al mercado laboral

El análisis de conglomerados permitió clasificar a los 71 departamentos sobre la base de los puntajes de los cinco factores principales. Los estratos de IDECV resultantes reconocieron afinidades entre departamentos de distintas provincias. En efecto, los estratos suelen tener una composición que atraviesa las fronteras provinciales. Sólo el Estrato II se compone en una elevada proporción por departamentos de una sola provincia (Mendoza). Opuestamente a éste, el Estrato IV se nutre por partes iguales por departamentos de La Rioja y Catamarca, mientras que el Estrato I lo conforman similares cantidades de departamentos de Mendoza, La Rioja y San Juan.

La validez de la estratificación de departamentos se probó mediante la asociación con un indicador externo: el '% de hogares con NBI'. De acuerdo a este examen, se halló que la variación de la incidencia del NBI era más explicada por el Estrato de IDECV que por la provincia de pertenencia del departamento.

Finalmente resta hacer un comentario acerca de la conformación de los estratos (18). El reconocimiento de afinidades entre las áreas intraprovinciales debería ser confirmado por nuevos hallazgos y complementado con un análisis geográfico que de cuenta de la continuidad espacial de estas áreas. Entre tanto, con los resultados disponibles, se advierte que en el Estrato I (con menor incidencia de pobreza) se reúnen los distritos capitales de las cuatro provincias. Por otra parte, el estrato con las condiciones de vida más deficitarias (Estrato IV) sólo se compone por departamentos de las provincias de La Rioja y Catamarca. Estos dos señalamientos demuestran el sentido del planteo que orienta este trabajo, ya que sin desconocer la desigualdad existente entre las provincias de la Argentina, también sería posible hallar que ciertas realidades se reproducen atravesando los límites provinciales.

Notas

(1) Los criterios de selección de los indicadores que dan cuenta de los diferentes comportamientos de las respectivas estructuras productivas son: a) Presencia dominante del sector público tanto como eje del proceso de acumulación o como principal demandante de fuerza de trabajo. b) Significativo rol de la actividad industrial desarrollada, con preferencia en etapas medias de acumulación. c) Estructuras productivas ubicadas en centros urbanos no coincidentes con la sede del poder político nacional o provincial, donde el empleo público posee muy limitada incidencia. d) Actividades productivas diversificadas, con fuerte presencia de procesos de transformación modernos. e) Estructuras económicas estrechamente ligadas a procesos extractivos, ya sean de origen minero o agrícola, ubicadas en las áreas de influencia de las aglomeraciones respectivas e inciden en el comportamiento productivo y laboral de las mismas.
(2) Ver anexo con las definiciones de los indicadores utilizados.
(3) García Ferrando (1985), p. 432
(4) Ferrán Aranaz (1996) Capítulo17
(5) La matriz completa de datos se componía de veintiséis indicadores. Para el análisis factorial se excluyó el '% de hogares con NBI' que fue posteriormente utilizado para la validación externa.
(6) La nómina completa de indicadores analizados se consigna en el cuadro 1 del Anexo Estadístico.
(7) La matriz de correlación se encuentra en el cuadro 2 y la aplicación de una prueba de significación a las respectivas correlaciones se dispone en el cuadro 3 del Anexo Estadístico.
(8) Para mayores detalles sobre el análisis factorial de componentes principales se puede consultar a Comrey, A.(1985).
(9) Los factores que se eligieron fueron aquellos que tenían la mayor correlación lineal con las variables. El criterio estadístico que se tomó fue seleccionar los factores que tuviesen un 'eigenvalue' superior a 1, es decir aquellos que tienen la mayor proporción de varianza explicada.
(10) Los resultados de la extracción de factores por componentes principales figuran en el cuadro 4 del Anexo Estadístico.
(11) El método que se aplicó fue el Varimax.
(12) La correlación de cada indicador con el respectivo factor se encuentra en el cuadro 5 del Anexo Estadístico.
(13) Ejercicios adicionales, que exceden el marco de esta presentación, confirmaron que los resultados obtenidos con seis factores eran menos significativos que los desarrollados con cinco.
(14) Ferrán Aranaz, M. (1996) Cap. 17

(15) El detalle del procedimiento de cálculo de los centroides se presenta en el Cuadro 12 del Anexo Estadístico.
(16) La descripción de las puntuaciones medias de cada estrato según factor se halla en el cuadro 8 del Anexo Estadístico.
(17) La definición del índice de NBI se encuentra en INDEC (1984).
(18) La nómina completa de los 71 departamentos clasificados según el estrato de IDECV, se presenta en el Cuadro 1 del Anexo estadístico.

Bibliografía

Comrey, A. (1985) *Manual de análisis factorial.* Cátedra. Madrid

Cortés, F. y Rubalcava, R. M. (1991) *Consideraciones sobre el uso de la estadística en las ciencias sociales: estar a la moda o pensar un poco.* El Colegio de México (mimeo). México.

Ferrán Aranaz, M. (1996) *SPSS para Windows. Programación y análisis estadístico*, Serie Mac Graw-Hill de Informática, Mac Graw-Hill – Interamericana. Madrid.

García Ferrando, M. (1985) *Socioestadística. Introducción a la estadística en sociología.* Alianza. Madrid.

Heintz, P. (1970) *Un paradigma sociológico del desarrollo. Con especial referencia a América Latina.* Del Instituto. Buenos Aires.

INDEC (1994) *La Pobreza en Argentina*. Buenos Aires.

INDEC (1996) *Situación y evolución social provincial. Mendoza*. Buenos Aires.

INDEC (1997) *Situación y evolución social provincial. San Juan*, Buenos Aires.

INDEC (1997) *Situación y evolución social provincial. Catamarca.* Buenos Aires.

INDEC (1998) *Situación y evolución social provincial. La Rioja.* Buenos Aires.

Kinneary, T. y Taylor, J. R. (1990) *Investigación de Mercados. Un enfoque aplicado.* Mac Graw-Hill. Madrid.

Manzanal, M. (1995) "Globalización y ajuste en la realidad regional argentina: ¿reestructuración o difusión de la pobreza?" en *Realidad Económica*, N° 134. pp 67-82. Buenos Aires.

Mayo, A. (1995) "Plan Cavallo y Economías Regionales: el mito de la salida exportadora" en *Desarrollo Económico*, N° 135 pp. 10 –34. Buenos Aires.

Minujin, A. (comp) (1992) *Cuesta Abajo. Los nuevos pobres: efectos de la crisis en la sociedad argentina.* UNICEF/LOSADA. Buenos Aires.

Rofman, A. (1995) "Las economías regionales: un proceso de decadencia estructural" en Pablo Bustos (comp), *Más allá de la Estabilidad,*Fundación Friedrich Ebert (pp 159-189). Buenos Aires.

Rofman, A. (1974) *Desigualdades regionales y concentración económica. El caso argentino.* SIAP-PLANTEOS. Buenos Aires.

Rofman, A., Manzanal, M. y otros (1987) *Políticas Estatales y Desarrollo regional. La experiencia del Gobierno Militar en la región del NEA (1976-1081).* CEUR. Buenos Aires.

Secretaría de Programación Económica (1994) *Exportaciones Provinciales y regionales. Período 1988-93*. Buenos Aires.

Vaca, A. y Cao, H. (1998) "Endeudamiento Provincial" en *Desarrollo Económico*, N° 158 (pp. 135 –147). Buenos Aires.

Rofman, A. (1997) *Convertibilidad y desocupación en la Argentina de los 90. Análisis de una relación inseparable.* Ed. Oficina de Publicaciones del CBC, Colección CEA – CBC, UBA. Buenos Aires.