

Una metodología alternativa para la construcción de agregaciones socioespaciales homogéneas

Mg. Juan José Primosich (Universidad Nacional de Tres de Febrero)

jprimosich@untref.edu.ar

1. Objetivo del trabajo

El objeto de este trabajo es analizar el modo en que se realizan estas agregaciones mediante el Análisis de Componentes Principales en Dos Etapas, utilizando como unidad de análisis la información de la Encuesta Permanente de Hogares publicada por el INDEC. Al aplicar esta técnica se buscan realizar dos tipos de aportes. Por un lado, desde un punto de vista procedimental, al aplicarlo se pueden delimitar agregaciones socioespaciales homogéneas en base a un criterio de interés específico. Por el otro, la aplicación en forma crítica de este método expone el modo en que las decisiones tomadas por el investigador antes y durante la investigación afectan los resultados obtenidos.

2. Antecedentes

Para intentar medir con precisión la privación, se utilizan datos de encuestas o censos con el objetivo de clasificar a las poblaciones o áreas de riesgo, que permitan a los *policy decision-makers* formular estrategias que permitan asignar una combinación de recursos sociales y materiales para mejorar esta situación. En este sentido, resulta siempre necesario relevar y desarrollar las herramientas que facilitan explorar las dimensiones económicas y sociales y que permitan establecer criterios de asignación de los recursos disponibles en una sociedad y que a su vez permitan mapear las necesidades en cada área geográfica individual. De este modo se puede medir y representar las condiciones de privación en un territorio definido.

Es necesario tener siempre presente que la objetivización de la realidad deriva en una tecnificación que aborda el cómo y no el porqué del objeto de estudio (Pinto Verdugo, 2007). De esta manera, el apriorismo lógico-matemático en la construcción de teoría — que en este caso pertenecen al Análisis Factorial y/o al Análisis de Componentes Principales — posee la gran desventaja de que sus resultados sólo son directamente extrapolables a aspectos descriptivos de la estructura analizada no pudiendo entregar respuestas que expliquen el porqué de ese ordenamiento espacial. En este tipo de construcción prevalece la metodología matemática y de ahí la necesidad de la rigurosidad excesiva en sus análisis numéricos, pero

no explicitan en los mismos parámetros de especificidad la relación entre lo experiencial y la explicación, que es expuesta en términos discursivos.

Profundizando en los aspectos del conocimiento y del uso de la técnica, ya se ha reflexionado críticamente sobre la naturaleza y las consecuencias de este modo de llevar este tipo de análisis (Jackson & Borgatta, 1981; Marradi, 1978; Tullio-Altan & Marradi, 1976). En este sentido, es importante recordar que el vínculo de la ortogonalidad entre los componentes o los factores extraídos, cualquiera sea su número, tienen sólo motivos matemáticos y es resultado del procedimiento de extracción. Sin embargo, en la investigación empírica en ciencias sociales raramente se utilizan dimensiones conceptualmente independientes entre sí. Entonces puede darse el caso que algunas variables están asociadas más o menos en la misma medida con más de un componente: la ortogonalidad es artificialmente impuesta por una exigencia técnica (Marradi, 1981). El segundo inconveniente que encontramos es el intento de construir un índice asignando puntajes a los casos de estudios, teniendo en cuenta todo el conjunto de variables sometidas inicialmente al análisis, inclusive las que han manifestado tener saturaciones nulas o irrisorias sobre el componente o los factores en cuestión. El culto al fetiche de la objetividad induce a intervenir lo menos posible con decisiones como excluir determinada variable ya que esas decisiones son percibidas como subjetivas y arbitrarias, y por tanto no científicas. De este modo se pierde de vista que esa construcción de la objetividad se basa sobre una decisión inicial subjetiva, arbitraria e imposible de obviar: la composición de la canasta inicial de variables. La consecuencia práctica de este comportamiento 'objetivo' es que el índice se compone de coeficientes componenciales ínfimos, cuyo efecto en el puntaje final es insignificante. Esto tiene el defecto semántico de oscurecer la presencia de aquello que es verdaderamente esencial en la dimensión, dificultando su interpretación.

Usualmente los coeficientes de correlación puestos en una matriz no son calculados entre variables cardinales, sino entre variables generosamente bautizadas de intervalo a pesar de que no haya ninguna prueba de que los intervalos que dividen cada par de puntos adyacentes sobre la escala sean efectivamente iguales, violando la definición. Además, se encuentran coeficientes de correlaciones calculados entre variables dicotómicas e incluso entre variables categoriales no ordenadas. Pero los coeficientes de correlación entre variables dicotómicas resultan gravemente distorsionados si las frecuencias no están equilibradas (Marradi, 1997). La idea de aplicar la técnica que presupone variables cardinales a variables cuyas categorías no están ni siquiera ordenadas, testimonia las aberraciones inducidas por la manía de cuantificar (ver ejemplos en Rummel, 1967).

La aplicación del Análisis de Componentes Principales en Dos Pasos se revela útil cuando queremos descubrir dimensiones conceptuales de gran generalidad a partir de muchas variables y para una gran cantidad de casos como en este ejemplo.

3. Método y fuentes de información

Para este ejercicio se utilizaron los resultados de la Encuesta Permanente de Hogares (EPH) correspondientes al tercer trimestre de 2017 y como método el Análisis de Componentes Principales en Dos etapas (ACP) concebido por Alberto Marradi para descubrir las dimensiones socioeconómicas de la privación.

4. Procedimiento

Como primer paso se preparó la matriz de datos sobre la cual se trabajaría. Se llevó a cabo una primera selección de las variables que se someterían al análisis de componentes principales. Las dimensiones que interesan al análisis pueden ser múltiples y cada una podría ser representada por muchas variables. Sería posible resolver el problema escogiendo arbitrariamente una o dos variables que representen alguna dimensión. Pero podría suceder que dos dominios de estudio sean similares respecto a alguna variable, pero difieran mucho en otra. Por esto, es preferible combinar la información provista por una gama de variables que sean al mismo tiempo empíricamente definidas y cuantificadas y semánticamente interpretadas. La selección y/o construcción de las variables iniciales tiene que responder más a la curiosidad del investigador que a la disponibilidad de las mismas. En este sentido, explicitar cuál es el criterio para elegir unas variables (y descartar otras) para la construcción de la matriz de datos original agrega valor al procedimiento. Construir las variables iniciales supone realizar reelaboraciones simples (en general, transformando cada dato bruto en una proporción respecto a una base oportunamente elegida) y a veces cálculos un poco más complejos.

Para la primera etapa se utilizaron las 14 variables — indicadores indirectos de privación, ya que en sí mismos no implican privación — que fueron seleccionadas para el estudio del Dr. Ivaldi (2020).

Esta variante en Dos Pasos fue concebida en 1978 por el profesor Alberto Marradi y publicada en revistas italianas, inglesas y españolas. Metodológicamente, la ventaja de esta estrategia es el control conceptual de cada uno de los pasos del procedimiento teniendo en cuenta que la principal preocupación del investigador debe ser la utilización de técnicas que le permitan obtener resultados que cumplan con sus objetivos cognitivos.

Esta propuesta busca superar la rutina de los programas estadísticos y tomar una posición activa considerando las afirmaciones de Thurstone (1947) cuando destacó que lo teóricamente importante en este tipo de análisis ocurre en la mente del investigador antes del análisis, cuando los indicadores se seleccionan y, después del proceso, cuando se interpretan los factores.

La primera etapa tiene un objetivo exploratorio; trata de reducir el número de variables para facilitar el control intelectual sobre toda la situación y la interpretabilidad de las soluciones propuestas.

Para investigar el número y naturaleza de estas dimensiones, el conjunto de las variables elegidas se somete a un ACP y sus distancias se representan en diferentes diagramas bidimensionales en el espacio de las dimensiones $v-1$, donde la conexión de ortogonalidad es matemáticamente necesaria para poder configurar el espacio de componentes o factores. En aquellas gráficas donde se muestran diferentes factores (factor 1 contra factor 2, factor 1 contra factor 3, factor 2 contra factor 3, etc.) es sólo un paso intermedio que sirve para representar la configuración de los vectores, reconocer las dimensiones e identificar los subconjuntos semánticamente homogéneos de variables dentro del conjunto inicial. Los subconjuntos dispuestos en esos diagramas muestran grupos de variables claramente separados. Estos clústeres son los que necesitan atención: cuando existen, hay algo que interpretar.

¿Por qué se hace hincapié en que sean semánticamente relevantes? Porque con el análisis de componentes principales realizado de la forma habitual, se obtienen factores como resultado de aspectos matemáticos sin considerar su interpretabilidad: sucede a menudo que los factores emergentes son matemáticamente perfectos, pero conceptualmente inútiles.

El primer paso finaliza cuando de la canasta original de variables surgen subconjuntos de variables para cada dimensión individualizada, pudiendo las variables originales estar en más de un subconjunto. Las variables que componen cada uno de estos subconjuntos semánticamente homogéneos constituirá la cesta de variables sobre la que se realizará el procedimiento de refinamiento en el siguiente paso.

Lo único que realmente importa es la posición respectiva de los puntos que representan las variables en el espacio multidimensional. Una vez logrado esto, el sistema de ejes ha hecho su trabajo: lejos de cosificar se los puede olvidar. Como en la fotografía de una habitación, las diferentes posiciones en las que se puede colocar la lente son arbitrarias y sin sentido: solo se utilizan para determinar las posiciones relativas de los objetos en la habitación.

Este procedimiento puede ser criticado por ir más allá de los aspectos meramente objetivos. Pero esto es inevitable porque las restricciones que impone la ortogonalidad suelen ocultar el mensaje semántico presente en los datos.

La ventaja más importante desde el punto de vista metodológico de la estrategia de dos pasos es dar al investigador un control conceptual total de cada uno de los pasos del procedimiento. El investigador debe usar esto, como cualquier otra técnica de análisis de datos, para obtener resultados que satisfagan sus preguntas.

El segundo paso consistirá en el refinamiento por separado de cada dimensión, sometiendo cada subconjunto a un ACP hasta obtener una dimensión satisfactoria, teniendo en cuenta no sólo los pesos de cada variable en los ejes cartesianos sino también su grado de distancia con las demás. Se puede considerar la posibilidad de eliminar algunas variables del grupo. Así, se procede como de costumbre.

Si una variable parece alejarse del núcleo de las demás, es conveniente excluirla y volver a someter el subconjunto restante a la técnica. Cuando las variables que permanecen en el gráfico son cercanas entre sí y semánticamente importantes, el proceso de refinamiento puede considerarse suficiente. Esto favorece la claridad semántica de cada dimensión.

El siguiente paso es elegir las variables en base a 2 criterios: uno semántico y otro espacial (la posición de la variable en el gráfico viene determinada por el peso que tuvo sobre las dos dimensiones que forman los ejes de la representación).

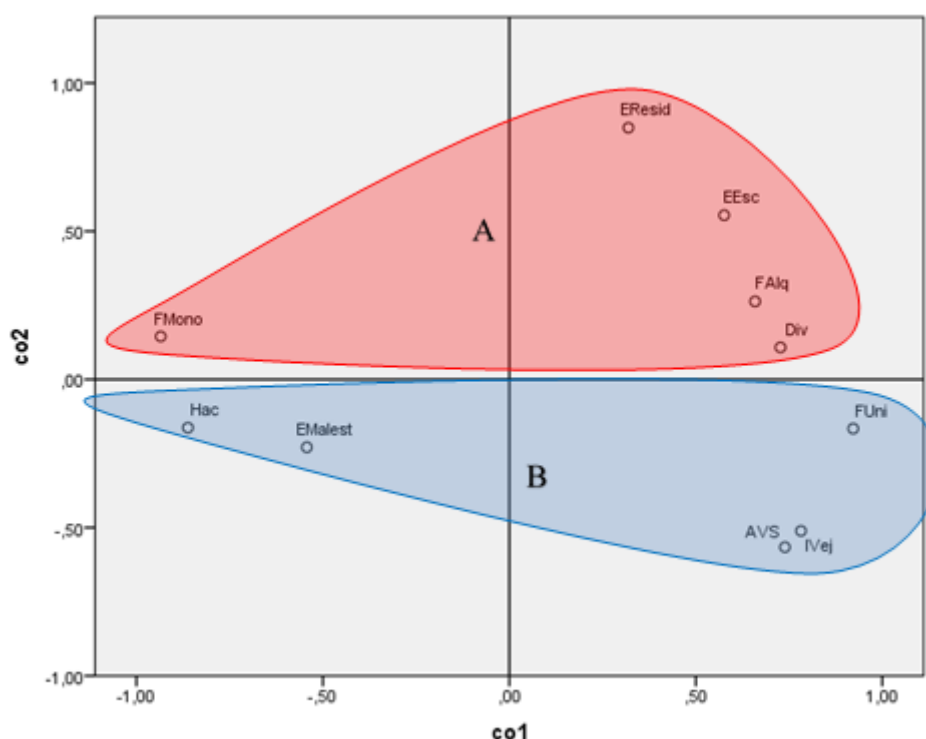
En esta fase, las decisiones se toman básicamente en base al criterio semántico, teniendo en cuenta los posibles solapamientos.

En el caso del ejemplo que se trata en este trabajo, después de haber analizado en profundidad los diferentes pares de componentes que surgieron del primer ACP, podemos identificar, tanto con criterios espaciales como semánticos, dos dimensiones: una que se encuentra entre el primer y segundo cuadrante, que llamaremos dimensión A, y una segunda que se encuentra en el tercer y cuarto cuadrante, al que llamaremos B.

Pesos de las variables en la dimensión PRIVMAT-SOCIOECO

Matriz de componentes

	Componente
	1
Hacinamiento	-,821
Índice vejez	,884
Ancianos viven solos	,852
Edificio en mal estado	-,547
Familia unipersonal	,924



A la izquierda podemos ver variables que indican los rasgos asociados a la privación material: familias que viven en viviendas hacinadas y edificios en mal estado. Del lado derecho es posible observar variables relacionadas con la carencia socioeconómica: número promedio de miembros por familia, presencia de ancianos que viven solos, % de la población igual o mayor a 65 años.

Esta solución se puede llamar PRIVMAT-SOCIOECO = privación material / privación económica.

Los coeficientes enumerados en la tabla se utilizan para construir la fórmula mediante la cual se calcula la puntuación de cada caso en la dimensión encontrada. A partir de la

interpretación de los principales factores y del refinamiento de la dimensión (en este caso, solo uno), es posible atribuir a cada conglomerado un puntaje componente en la dimensión.

De esta forma, es posible enriquecer la matriz de datos con un nuevo vector, que permite un control empírico de las decisiones tomadas en las fases posteriores del análisis.

Por lo tanto, el resultado de la comprobación empírica es claro: para quienes tienen un mínimo conocimiento sobre la distribución de los recursos en las provincias argentinas, las cifras hablan por sí solas. Y lo que también es importante es que estos resultados coinciden con los obtenidos en el estudio mencionado al principio.

Puntuaciones de los conglomerados sobre la dimensión PRIVMAT-SOCIOECO

AGLOMERADO	Índice
Ciudad de Bs As	-0,072
Gran Paraná	-0,246
Gran Rosario	-0,276
Bahía Blanca - Cerri	-0,271
Mar del Plata - Batán	-0,284
.....
Gran Tucumán - T. Viejo	-0,565
Gran Resistencia	-0,583
Gran San Juan	-0,583
Gran Catamarca	-0,603
Salta	-0,620
S. del Estero - La Banda	-0,699

Bibliografía

Ivaldi, E., Parra Saiani, P., Primosich, J. J., & Bruzzi, C. (2020). Health and Deprivation: A New Approach Applied to 32 Argentinian Urban Areas. *Social Indicators Research*. <https://doi.org/10.1007/s11205-020-02369-w>

Jackson, D. J., & Borgatta, E. F. (1981). *Factor analysis and measurement in sociological research : a multi-dimensional perspective* (Sage (ed.)). Sage.

Madge, N., & Brown, M. (1982). Despite the welfare state : a report on the SSRC/DHSS programme of research into transmitted deprivation. In N. Madge (Ed.), *Studies in deprivation and disadvantage*. Heinemann Educational.

Marradi, A. (1978). *Die Faktorenanalyse und ihre Rolle in der Entwicklung und Verfeinerung empirisch nützlicher Konzepte: Vol. XXX* (pp. 488–513).

Marradi, A. (1981). Misurazione e scale: qualche riflessione e una proposta. *Quaderni Di Sociologia*.

https://www.academia.edu/3695998/Misurazione_e_scale_qualche_riflessione_e_una_proposta

Marradi, A. (1997). *Linee guida per l'analisi bivariata dei dati nelle scienze sociali*. Angeli.

http://books.google.it/books?id=hpF_IKICyJQC&printsec=frontcover&dq=marradi+analisi+bivariata&hl=it&ei=4OJcTsWjHsOWOqbhtPQC&sa=X&oi=book_result&ct=result&resnum=1&ved=0CCsQ6AEwAA#v=onepage&q&f=false

Pinto Verdugo, M. (2007). *Los modelos explicativos de la estructura urbana : una visión crítica de su desarrollo al interior de la consolidación de la racionalidad moderna*. <https://repositorio.uchile.cl/handle/2250/100994>

Rummel, R. J. (1967). Understanding factor analysis. *Journal of Conflict Resolution*, 11(4), 444–480. <https://doi.org/10.1177/002200276701100405>

Spicker, P. (2009). The Nature of a Public Service. *International Journal of Public Administration*, 32(11), 970–991.

Tullio-Altan, C., & Marradi, A. (1976). Valori, classi sociali, scelte politiche : indagine sulla gioventù degli anni Settanta. In *Nuovi saggi italiani 20*. Bompiani.