El índice de marginación desde otras perspectivas metodológicas

Yolanda Téllez Vázquez[1]

Resumen

En sus inicios, la necesidad de estimar la marginación se centró en disponer de una medida que permitiera conocer el impacto de las carencias socioeconómicas en cada unidad geográfica del país, pero sin tomar en cuenta la variable tiempo. Es por ello que en este trabajo se revisan algunas propuestas sugeridas en los últimos años para que la estimación del índice y grado de marginación pueda subsanar esta limitante (se sabe que la pérdida de discriminación de los indicadores y las restricciones que le impone la aplicación de la técnica de estratificación actual son otros aspectos a considerar). En el documento se realizan varios ejercicios aplicados al año 2010 con los diversos métodos alternativos. Los resultados obtenidos concluyen que ningún procedimiento es completamente consistente, aunque la aplicación de conjuntos difusos resulta uno de los mejores acercamientos metodológicos.

Términos clave: marginación, actualización, métodos alternativos de cálculo.

Introducción

La marginación es un fenómeno estructural y multidimensional, en el cual intervienen factores asociados directamente a la población y a sus viviendas, esta misma naturaleza vuelve compleja su medición y análisis.[2] Su connotación territorial la diferencia claramente de la pobreza, cuya medición está asociada a la persona; la condición de pobreza es de la persona, mientras que el concepto de marginación caracteriza a aquellos grupos que han quedado al margen de los beneficios del desarrollo nacional y de la riqueza generada, pero no necesariamente al margen de la generación de esa riqueza ni mucho menos de las condiciones que la hacen posible (COPLAMAR, 1977).

La relevancia del índice reside en que se ha convertido en un indicador que da cuenta de la distribución espacial de la marginación en el territorio nacional, mostrando en primer plano el panorama estatal y municipal. Posteriormente, en 1995, se agregó la estimación de localidad y en 2000, la de áreas geoestadísticas básicas (AGEB) urbanas. Aunque el presente análisis se centra en la principal crítica realizada al cálculo de la marginación, los siguiente aspectos también han estado presentes:

- La no comparabilidad en el tiempo y entre distintas unidades geográficas, cuestión relacionada con el empleo del método de Análisis de Componentes Principales (ACP),

1 Se agradecen las aportaciones del Act. Jorge López Ramírez.
2 Antes de la aparición de la primera publicación del índice de marginación en 1990 se habían realizado diversos ejercicios, cuyo resultado final fue el empleo de la metodología actual.

- El considerar que alguno de los indicadores deja de ser útil en términos de su factor discriminante,
- Los cambios en los estratos, debido al empleo de la estratificación óptima de Dalenius y Hodges, que es una técnica univariada aplicada a un fenómeno multidimensional.

En este trabajo se atenderá la comparabilidad en el tiempo. Al respecto, es de suma importancia mencionar que, en su momento, se privilegió la necesidad de contar con un estadístico asequible que automatizara la estimación, siendo consistente con el propósito de reducir la multidimensionalidad del fenómeno. No obstante, no se previó el uso generalizado que se le daría para fines de políticas públicas y la necesidad de su evaluación y de conocer los cambios a través del tiempo.

Por tanto, el objetivo del artículo es dar a conocer otras propuestas metodológicas que subrayan la principal limitante de la actual forma de cálculo del índice de marginación, la comparabilidad en el tiempo. Dichas propuestas corresponden a algunas de las exposiciones llevadas a cabo en el "Seminario de actualización del marco conceptual y metodológico del índice de marginación", realizado en Ixtapan de la Sal, Estado de México, del 20 al 22 de junio de 2012.

El documento está organizado en cuatro apartados: el primero, proporciona el panorama de la marginación en el contexto actual; el siguiente, expone diversas metodologías propuestas; después se muestran los resultados de la aplicación de éstas técnicas manteniendo los supuestos actuales; y, finalmente, se expresan algunas consideraciones, resultado del análisis de las propuestas presentadas.

Panorama de la marginación

Estimar la marginación ha sido una acción de gran relevancia, por lo que desde 1980 se estableció la necesidad de caracterizarla de tal forma que fuera una medida capaz de proporcionar información sobre el comportamiento del fenómeno. Desde sus inicios, se planteó un esquema que consideraba la situación económica, política y social de las áreas geográficas marginadas, a fin de dotarlas de elementos materiales y de organización suficientes para participar de forma más equitativa del disfrute de los beneficios del desarrollo.

Con el propósito de construir una medida que diera cuenta de la intensidad del fenómeno, se buscó generar un indicador que evaluara el impacto global de los déficits y que cumpliera con ciertas características, facilitando el análisis territorial, reduciendo la dimensión original, reteniendo el máximo posible de información, así como estableciendo las relaciones entre sus indicadores, y ordenando las unidades geográficas.

Para su estimación, como fuente de información se consideraron los censos y conteos de población y vivienda, debido a que cuentan con la cobertura, nivel de desagregación y confiabilidad que se requieren para la construcción del índice. La estimación del índice inicialmente se realizó en dos niveles geográficos: entidades federativas y municipios; posteriormente, se agregaron las localidades, y para el siguiente, las AGEB[3] urbanas; en estos dos últimos niveles fueron precisos algunos cambios en las variables. En su construcción, el índice considera dimensiones estructurales como: educación, vivienda, salud, ingresos, dispersión de la población o disponibilidad de bienes.[4]

Dados los requerimientos para el cálculo del índice de marginación, se utilizó el método matemático de Análisis de Componentes Principales (Harman, 1976), que a través de un conjunto de variables más pequeño pretende interpretar de forma sencilla el fenómeno original.[5] Para este cálculo se puede emplear tanto la matriz de covarianzas como la de correlaciones. La primera, se emplea cuando las variables originales tienen aproximadamente la misma varianza, de forma que el

[3] Área Geoestadística Básica o AGEB es la subdivisión de los municipios o delegaciones que conforman el país, integrando las unidades primarias de muestreo y la organización de la información estadística. Tiene tres atributos fundamentales: a) es perfectamente reconocible en el terreno por estar delimitada por rasgos topográficos identificables y perdurables; b) por lo general es homogénea en cuanto a sus características geográficas, económicas y sociales; c) su extensión es tal que puede ser recorrida por una sola persona. Las AGEB se clasifican en más y menos urbanizadas, dependiendo de su densidad de viviendas.

[4] Las dimensiones de ingresos y dispersión solo se consideran en el cálculo estatal y municipal, la de salud, únicamente en el caso de AGEB urbana. Las dimensiones de educación y vivienda se consideran en todas las unidades con algunas variantes.

[5] Para ello se requiere que las variables originales presenten cierto nivel de correlación, lo cual es evidente en el caso de los indicadores utilizados para el cálculo de la marginación.

cálculo de las componentes se hace en términos de las variables originales. La segunda, se utiliza cuando las escalas de medición de las variables son diferentes o sus varianzas se encuentran muy alejadas. En este último caso, las componentes principales se obtienen de la estandarización de las variables originales, siendo ésta la matriz que se ha empleado para obtener los índices de marginación.

Dada su construcción, los valores de los indicadores oscilan entre cero y cien, señalando la ausencia o presencia del déficit, respectivamente. A fin de eliminar los efectos de escala entre las variables,[6] éstas se estandarizan con el promedio aritmético y la desviación estándar de cada indicador. Para su cálculo, se emplea el paquete estadístico Statistic Package for Social Sciencies (SPSS), que proporciona Componentes Principales estandarizados, con media cero y desviación estándar uno. Los coeficientes de ponderación se dan recalculados para las variables originales.

El propósito de aplicar el método de componentes principales no es obtener el menor número posible de componentes que expliquen una parte significativa de la varianza (objetivo usual en las aplicaciones del Análisis de Componentes Principales), sino proyectar el espacio definido por los nueve indicadores sobre uno unidimensional, de tipo escala de intervalo, mediante el cual es posible establecer un orden absoluto entre las unidades geográficas. Por ello, solo se considera el primer componente del análisis; además, dadas las características de la metodología aplicada, este componente es el que devuelve mayor información del fenómeno.

Posterior a la aplicación del método de componentes principales se obtienen los grados de marginación. La aplicación de la técnica de estratificación óptima de Dalenius y Hodges (Dalenius, 1959) asegura que los estratos formados tengan varianza mínima al interior y máxima entre cada uno de ellos, es decir, se forman estratos homogéneos al interior y heterogéneos entre ellos. Para su interpretación, se establecen cinco estratos o grados que muestran la intensidad de la marginación en las unidades geográficas: muy alto, alto, medio, bajo y muy bajo.

En cada nivel geográfico se efectúa el mismo procedimiento para calcular el índice y establecer los rangos de agrupación. Debido a que tanto el análisis de componentes como la técnica de estratificación empleadas se calculan con la variabilidad de cada categoría geográfica y para el año censal correspondiente, no es posible hacer comparaciones, ni entre niveles, ni en un lapso determinado, solo es posible equiparar entre unidades de un mismo nivel y para el año establecido. Por tal motivo, es necesario que se revisen diferentes propuestas metodológicas.

Metodologías propuestas

En los últimos años, la metodología empleada por el Consejo Nacional de Población (CONAPO) para calcular el índice de marginación ha sido objeto de críticas de diversa índole, debido a la forma de medir la intensidad de la marginación. Los métodos que se presentan a continuación derivan de alternativas propuestas por varios autores, las cuales compensan las limitantes señaladas en el cálculo del índice de marginación.

Media aritmética (IAM)

El CONAPO realiza una estimación de la marginación que permite la comparación en el tiempo mediante una media aritmética o promedio simple, denominado índice absoluto de marginación (IAM). Este índice surgió como respuesta a la necesidad de conocer la evolución de la marginación a través de una medida sintética con la cual fuera posible comparar, entre unidades geográficas y de forma temporal, los avances o retrocesos en materia de marginación (CONAPO, 2004).[7]

$$IAM_j = \frac{\sum_i^9 x_i}{9}$$

[6] Aunque el recorrido de las nueve variables está acotado por la izquierda y la derecha, es necesario transformar las variables de tal manera que aquellas con una mayor varianza no predominen en la determinación del índice y vuelvan inoperante el análisis multivariado.

[7] Es importante considerar que las correlaciones entre los indicadores empleados son significativas.

Donde:

IAM_j índice absoluto de marginación para la unidad geográfica j.
x_i indicador socioeconómico para medir la intensidad de la exclusión i.

Es importante señalar que el primer análisis presentado por el CONAPO se realizó entre 1990 y 2000, a fin de comparar los índices y evaluar los cambios en la marginación en ese periodo. Para 2010, el ejercicio se realizó comparándolo con 2000. Para evitar fluctuaciones se fijaron las ponderaciones empleadas en cada uno de los indicadores. El CONAPO elabora este índice como complemento de la estimación del índice de marginación calculado a partir del método de componentes principales.

Este método reporta beneficios, entre ellos: no requiere de un punto de referencia en particular en el tiempo; su estimación resulta sencilla; los cambios tienen una interpretación directa y absoluta porque señalan qué tan alejada se encuentra cada unidad geográfica de análisis de la situación óptima; permite valorar el aumento o la reducción de la marginación al observar directamente los cambios en la reducción de los puntos porcentuales, independientemente de su valor inicial; también es posible apreciar las brechas existentes entre unidades geográficas. Con esta medida es posible comparar en el tiempo y entre unidades, sin embargo, no permite visualizar el peso de cada indicador, pues los nueve tienen la misma representatividad, y es sensible a valores extremos, que afectan su valor.

Media geométrica[8]

En la actualidad, esta metodología es aplicada internacionalmente para la medición del Índice de Desarrollo Humano (IDH); se estima mediante una media geométrica calculada a través de la raíz cúbica del producto de tres índices: salud o esperanza de vida, educación y riqueza o ingresos, de tal forma que el índice resultante es:

$$IDH = \sqrt[3]{IEV * IE * II}$$

Donde:

IDH índice de desarrollo humano.
IEV índice de salud o esperanza de vida, se calcula como, $IEV = \frac{EU - 20}{máxEu-20}$ ’ donde Eu corresponde a la esperanza de vida de la unidad calculada expresada en años.
IE índice de educación, se calcula como $IE = \frac{\sqrt{IAPE - IAEE}}{máx\sqrt{IAPE - IAEE}}$ ’ donde $IAPE$ es el índice de año promedio de escolaridad y $IAEE$ es el índice de años esperados de escolaridad.

Cada índice se estima de la siguiente manera:

$IAPE = \frac{APE}{máx\ APE - 0}$ ’ donde APE es el número de años promedio de escolaridad.

$IAEE = \frac{AEE}{máx\ AEE - 0}$ ’ donde AEE es el número de años esperados de escolaridad.

II índice de riqueza o ingreso, se calcula como $II = \frac{In\ (GNIpc) - In\ (100)}{In\ (40\ 000) - In\ (100)}$ ’ donde $GNIpc$ es el índice de PIB per cápita.

La ventaja de este método radica en que todos los valores pueden ser comparables si se mantienen los mismos máximos y es menos sensible a los valores extremos. Pero la interpretación de este método tiene un razonamiento menos directo en cuanto a su significado estadístico e interpretación, y su cálculo requiere establecer cotas, lo cual resulta difícil en el caso de la marginación. También debe considerarse que es posible que se indeterminen los cálculos si existen ceros en alguna variable,[9] lo que puede suceder en algunas de ellas dado el uso de porcentajes.

[8] Método expuesto por el Mtro. Rodolfo de la Torre, Director de la Oficina de Desarrollo Humano, Programa de las Naciones Unidas para el Desarrollo en México.

[9] También es posible que se indetermine si existen valores negativos.

Distancias euclidianas (índice de marginación euclidiano)[10]

El empleo del índice de marginación euclidiano (IME) sugiere recuperar, por un lado, las ventajas del índice de marginación del CONAPO, y, por otro, resolver los problemas de la comparación en el tiempo (González y Torres, 2011). Considera cada indicador como una dimensión, obteniendo un espacio de nueve dimensiones, llamado espacio métrico de la marginación, donde se establece el IME; por su construcción, los valores de cada unidad van de cero a 300. El IME es calculado como la distancia del punto $x=(x_1, x_2, ..., x_9)$ al punto CAM^9, llamado cero absoluto.[11]

$$IME_i = \sqrt{(Z_{i1} - 0)^2 + (Z_{i2} - 0)^2 + ... + (Z_{i9} - 0)^2}$$

Donde:

Z_{ij} es el valor del indicador j (1,...9) para la unidad geográfica i.

El nivel de marginación de una unidad dependerá de qué tan cerca o lejos se encuentre del CAM^9, señalando, con ello, lo menos o más marginado de cada unidad, respectivamente. Como la comparabilidad en el tiempo depende del cambio en sus indicadores, deberá asegurarse que las variables empleadas sean iguales.

Al tratarse de una distancia, la euclidiana es sensible a las unidades de medida de las variables; si se consideran las diferencias entre valores altos, éstos contribuirán en mayor medida que las diferencias entre valores bajos, así los cambios de escala determinarán cambios en la distancia entre unidades.[12] Si las variables empleadas están correlacionadas, la información será redundante, inflando la disimilaridad o divergencia entre las unidades.[13] La distancia euclidiana será, en consecuencia, recomendable cuando las variables sean homogéneas y estén medidas en unidades similares y/o cuando se desconozca la matriz de varianzas. Dado que el cálculo del índice de marginación ha demostrado la existencia de correlación entre las variables empleadas, es necesario ser cautelosos en su aplicación, pues éstas nos proporcionarán información, en gran medida, redundante. La propuesta determina un espacio llamado espacio euclidiano, donde delimita un mínimo y un máximo que permiten asegurar que la divergencia se presente.

Para la determinación de los estratos de marginación se propone una variante al método de las k medias al que refiere Neter y Wasserman (1974). Inician con $k=2$ estratos y la asignación de unidades por medio del método de k medias. Posteriormente, se calcula la medida de homogeneidad:

$$HO_k = \frac{\sum_{i=1}^{k} \sum_{j=1}^{n_i} (IMEij - \mu_i)^2}{n - k}$$

Donde:

IME_{ij} corresponde a la unidad geográfica j dentro del estrato i.

μ_i es la media en el estrato i.

n_i es el número de unidades en el estrato i.

Si la medida no es significativa comparada con un valor predeterminado, el proceso termina con $k=2$ estratos. En caso contrario, el proceso se repite con $k=3$ y continúa hasta encontrar el valor m, de tal forma que la medida de homogeneidad sea menor que el valor predeterminado.

10 Método expuesto por el Mtro. Javier González del Consejo Nacional de Población; trabajo realizado en colaboración con el Mtro. José Luis Torres.

11 El punto CAM^9=(0.0,...,0) es llamado cero absoluto de la marginación y señala que una unidad geográfica no presenta carencias en ninguna dimensión.

12 Es posible minimizar los cambios utilizando la distancia euclidiana normalizada.

13 Si se pondera la contribución de cada par de variables con pesos inversamente proporcionales a las correlaciones, se podrá evitar la disimilaridad o divergencia.

Componentes principales mediante la matriz de covarianza[14]

Otra alternativa para medir el impacto global de las carencias consiste en aplicar el Análisis de Componentes Principales, pero empleando la matriz de covarianza en lugar de la matriz de correlaciones (Bustos, 2009). Este procedimiento evita la esferización[15] que se puede presentar al estandarizar los indicadores, ya que los datos están expresados en la misma escala, porcentajes;[16] también se asegura que así se muestran con más claridad las inequidades. El ejercicio propone emplear al menos las dos primeras componentes para tomar en cuenta más de la mitad de la varianza explicada y aprovechar información significativa. Conviene aclarar que la esferización mencionada no invalida las conclusiones del análisis empleando la matriz de correlaciones, en tanto que el empleo de los dos primeros factores cambia el objetivo del índice, pues requiere de dos dimensiones para su interpretación en lugar de la unidimensionalidad utilizada actualmente.

Es más común el uso del grado de marginación que del índice, por lo que se sugiere emplear un procedimiento multivariado de estratificación (k-medias) para clasificar las unidades con base en los valores de las componentes. Para ello, se parte de una clasificación inicial que considera solo la primera componente; la segunda clasificación integra tanto a la primera como la segunda; la tercera contempla las primeras tres componentes y así sucesivamente hasta llegar a la que incluye a todas las componentes principales (Bustos, 2011). El número de componentes a elegir será aquel cuya medida de homogeneidad, después de la estratificación, sea la más grande y está dada por:

$$\sum_{j=1}^{k} \frac{\sigma_i^2}{\sigma_j^2\,(S)}$$

[14] Presentado por el Dr. Alfredo Bustos del Instituto Nacional de Estadística y Geografía.

[15] La esferización significa que la variable x de la ecuación x=Wz es linealmente transformada a una variable V=Qx, tal que la matriz de covarianza de V sea unitaria, transformación que es siempre posible.

[16] Como consecuencia del problema de esferización no es posible establecer direcciones de máxima varianza en las componentes principales obtenidas, al mismo tiempo que se ocultan inequidades entre las unidades.

Donde,

$$\sigma_j^2\,(S) = \frac{1}{N(N-1)} \sum_{i=1}^{N} (I_{ij} - \bar{I}_j)^2,\; j = 1,\ldots,k$$

y

$$\sigma_j^2\,(S) = \sum_{l=1}^{h} \left({N_l}/{N} \right)^2 \sigma_j^2\,(S_l)$$

con

$$\sigma_j^2\,(S_l) = \frac{1}{N_l\,(N_l - 1)} \sum_{i_l=1}^{N_l} (I_{i_l jl} - \bar{I}_{jl})^2, \begin{cases} j = 1,\ldots,k; \\ l = 1,\ldots,h; \end{cases}$$

$$N_l$$

La medida, simple en su cálculo, ignora la presencia de correlación entre los índices. La varianza poblacional de cada índice dividida entre el tamaño de la población será la varianza del promedio y la suma ponderada de varianzas “de promedios” dentro de estratos para cada índice será la varianza promedio del estimador estratificado. La medida a maximizar puede ser expresada en términos de precisiones:

$$\sum_{i=1}^{k} \frac{\sigma_i^2}{\sigma_j^2\,(S)} = \sum_{i=1}^{k} \frac{{1}/{\sigma_i^2\,(S)}}{{1}/{\sigma_i^2}} = \sum_{i=1}^{k} \frac{P_i\,(S)}{P_i}$$

La mínima precisión se alcanza cuando el cálculo de la varianza no considera ninguna estratificación. Por lo anterior, el valor de la medida será siempre mayor o igual al número de indicadores k. Para su cálculo, la medida requiere solo de resultados de cualquier paquete estadístico comercial. Si para alguno de los índices la estratificación no resulta en homogeneidad al interior de los estratos, el denominador se parecerá al numerador.

Esta opción se emplea tomando en cuenta que todas las variables tienen la misma unidad de medida y se trata de destacar cada una de las variables en función de su grado de variabilidad. Es importante señalar que la ordenación de las unidades de observación no es directa y el empleo de más de dos componentes principales podría dificultar el resumen e interpretación de los indicadores originales.

Análisis Factorial Confirmatorio[17]

El trabajo propone la elaboración de un índice sensible a la evolución de la marginación a lo largo del tiempo, con la condición de que produzca resultados equivalentes a los del CONAPO. Para cumplir con ese propósito, se sugiere emplear un Análisis Factorial Confirmatorio que, en combinación con la invarianza factorial, permite generar un índice cuyos valores pueden ser comparables en el tiempo.

El Análisis Factorial Confirmatorio se usa para comprobar factores hipotéticos, imponiendo restricciones a los pesos de los factores. De tal modo que, para eliminar una de las p variables en un factor específico, se impone la restricción de que su peso debe ser igual a cero. El método se usa en combinación con el Análisis Factorial Exploratorio, el cual sirve para examinar la dimensión de un conjunto de variables y expresa la variación y covarianza de este conjunto en función de factores F_k con $k =1,...,m$. Se llama exploratorio porque busca, por medio de un proceso iterativo, la mejor representación de las variables originales, reduciendo el número de dimensiones (factores). En otros términos, si $X_{i,j}$ con $i =1,...,n; j =1,...,p,$ denota el valor de la variable j de la unidad geográfica i, se tiene que:

$$X_{i,1} = \alpha_1 + \lambda_{1,1} F_{i,1} + \lambda_{1,2} F_{i,2} + ... + \lambda_{1,m} F_{i,m} + \varepsilon_{i,1}$$

$$X_{i,2} = \alpha_2 + \lambda_{2,1} F_{i,1} + \lambda_{2,2} F_{i,2} + ... + \lambda_{2,m} F_{i,m} + \varepsilon_{i,2}$$

$$X_{i,p} = \alpha_p + \lambda_{p,1} F_{i,1} + \lambda_{p,2} F_{i,2} + ... + \lambda_{p,m} F_{i,m} + \varepsilon_{i,p}$$

Donde:

$m < p$, α_i son las intersecciones.
$\lambda_{i,k}$ son las cargas de los factores.
$F_{i,k}$ son los factores.
$\varepsilon_{i,j}$ son los residuos con media cero e independientes de los factores.

Para estimar los pesos factoriales, se utiliza el método de máxima verosimilitud a través de factorizaciones sucesivas sobre la matriz de covarianzas hasta obtener la máxima varianza de los factores, mientras que para conseguir la comparación de los índices en el tiempo se propone aplicar el método de invarianza factorial longitudinal, remitiendo a la referencia Bollen y Curran (2006). La invarianza factorial se calcula usando la propiedad de que las cargas se parezcan a lo largo del tiempo y de esta manera no pierden la comparabilidad de las mediciones. La invarianza factorial se modela al hacer que las cargas sean similares en cada medición. Esto permite modelar la media y la varianza en cada medición, permitiendo estimar el efecto del tiempo.

Una de las ventajas que ofrece el método de análisis factorial es que las variables empleadas se pueden manejar en su escala original. La otra ventaja, y la más importante, es que permite que los índices se puedan utilizar a lo largo del tiempo para valorar la medida en la que se ha avanzado en el esfuerzo por reducir la marginación.

Si bien es cierto que esta propuesta es atractiva, se debe considerar que la invarianza factorial calculada no sea del conjunto de referencia, pues al considerarlo imposibilitaría la comparabilidad; las cargas factoriales resultan solo similares no iguales, así que requieren ser evaluadas por éstas y otras precisiones que surjan. Para su estimación, y principalmente su interpretación, es necesario considerar la necesidad de conocimientos estadísticos avanzados.

Análisis multivariado mediante conjuntos difusos

El análisis multivariado mediante conjuntos difusos, empleado para medir el impacto de la marginación, se caracteriza por ser una medida multidimensional que se generaliza para variables categóricas múltiples (Morales-Ramos y Morales-Ramos, 2008); su forma de cálculo consiste en:

$$I_i = \frac{1}{9n_i} \sum_{j=1}^{9} \sum_{h=1}^{n_i} z_{hij}$$

[17] Expuesto por el Dr. Delfino Vargas de El Colegio de México; trabajo elaborado conjuntamente con el Dr. Fernando Cortés.

Donde:

I_i índice calculado para la unidad geográfica i.
n_i población en la unidad geográfica i.

$$Z_{hij} \begin{cases} 1, \text{ si el individuo } h \text{ en la unidad geográfica } i \\ \quad \text{carece de la variable } j \\ 0, \text{ en otro caso } f(x) \end{cases}$$

Éste es el caso unidimensional, donde I_i tomará el valor uno si cada variable señala carencia para los individuos, es decir, pertenece completamente al conjunto difuso; y el valor cero cuando no pertenece al conjunto en ningún grado.

Las variables que intervienen en la valuación del impacto de la marginación tienen distinto orden de importancia. Se considera el caso difuso multidimensional y se generaliza para variables de categorías múltiples; se ajusta de acuerdo al número de individuos al que ésta afecta, así las variables requieren de un ponderador. Se recomienda emplear el logaritmo de la proporción de individuos con carencia de acuerdo a cada variable, debido a que la función inversa de la proporción puede tomar valores muy bajos, afectando el cálculo del índice.

Con esto, el índice se calcula como:

$$I_i = \frac{\sum_{j=1}^{9} \sum_{h=1}^{n_i} z_{hij}\, \omega_j}{n_i \sum_{j=1}^{9} \omega_j}$$

Donde:

I_i índice calculado ponderado para la unidad geográfica i.
n_i población en la unidad geográfica i.
ω_j ponderador del indicador j.

$$Z_{hij} \begin{cases} 1, \text{ si el individuo } h \text{ en la unidad geográfica } i \\ \quad \text{carece de la variable } j \\ 0, \text{ en otro caso } f(x) \end{cases}$$

Conviene aclarar que esta metodología se propone para medir la pobreza y se presenta como una alternativa para el cálculo del índice de marginación, donde no solo resulta ser una medida resumen, sino que considera toda la información disponible en las variables que intervienen. Asimismo, permite obtener un índice no solo con escala de intervalo, sino comparable en el tiempo.

Aunque la teoría de conjuntos difusos ofrece muchas ventajas respecto a los métodos tradicionales, se debe reconocer que su aplicación requiere de cierto grado de interpretación para delimitar correctamente el vínculo entre la realidad y la teoría. Además, es preciso establecer sólidos criterios para el cálculo de los ponderadores, de forma que se obtengan resultados de acuerdo a los procedimientos determinados; también se debe tener en cuenta que se considera el logaritmo del inverso de la proporción de los individuos con carencia de acuerdo con cada variable, lo cual debe contemplarse al momento de la interpretación.

Modelo de curvas latentes de crecimiento (CLC)[18]

La aplicación de este modelo es aplicable para el nivel de municipios, debido a que su implementación requiere de un mayor número de casos para que tenga sentido. Este tipo de modelo se utiliza para explorar modelos de maduración y desarrollo de individuos en el tiempo (McArdle & Epstein, 1987), y en el contexto de modelos con ecuaciones estructurales con datos incompletos (McArdle, 1994). En la presentación del modelo CLC se incluye la intersección y la pendiente de un modelo lineal como variables latentes aleatorias. Ambas variables se modelan considerando una parte común y un componente aleatorio específico para cada municipio (Bollen & Curran, 2006).

Dentro del análisis de variables latentes se distingue el modelado de variables latentes continuas y categóricas. En el análisis de variables continuas se ubica el análisis factorial (exploratorio y confirmatorio; Mulaik, 2010), así como los modelos de ecuaciones estructurales (Bollen, 1989). Por ejemplo, si las curvas de crecimiento son lineales, se pueden agrupar en clases basándose en las intersecciones y pendientes. Esta idea se expresa en la siguiente ecuación:

[18] Presentado por el Dr. Delfino Vargas de El Colegio de México; trabajo elaborado conjuntamente con el Dr. Fernando Cortés.

$$P(c = k \mid t) = \frac{exp\ (\pi_{0,k} + \pi_{i,k}t)}{\sum_{i=1}^{K} exp\ (\pi_{0,k} + \pi_{i,k}t)}$$

que representa la probabilidad de que la trayectoria de los municipios pertenezca a la clase genérica K, donde k puede variar desde 1 hasta K. Esta formalización origina un modelo multinomial, que se estima a través de la ecuación:

$$log\ \frac{P(c = k \mid t)}{P(c = K \mid t)} = \pi_{0,1} + \pi_{1,1}t$$

Con el fin de revisar si la tendencia caracteriza a los municipios, se realiza un Análisis de Clases Latentes que, con base en el modelo de crecimiento, identifica y clasifica a los municipios considerando trayectorias similares y distinguiendo aquellos con trayectorias diferentes, usando la hipótesis multinomial para analizar la heterogeneidad de las trayectorias.

Este análisis se denomina así debido a que la variable latente es discreta. Una clase se caracteriza por un patrón de probabilidades condicionales que indican la probabilidad de que las variables tomen determinados valores. Según Goodman (1974), las clases se forman en función de una variable latente categórica que genera una división en clases latentes exhaustivas y mutuamente excluyentes, y en cada clase latente las variables observadas son estadísticamente independientes. También se puede utilizar para clasificar casos en función de su máxima verosimilitud de pertenencia a una clase.

El análisis de clases latentes se basa en el concepto de probabilidad y recurre a los datos examinados para estimar los parámetros del modelo: la probabilidad de cada clase, cuya suma debe ser igual a 1 y las probabilidades de respuesta condicional, lo cual representa la probabilidad de una respuesta particular en una variable observada condicionada por la pertenencia a una clase latente determinada.

$$f\ (y_i|\theta) = \sum_{k=1}^{K} \pi_k \prod_{j=1}^{J} f_k\ (y_{ij} \mid \theta_{jk})$$

Donde:

y_i representa las respuestas de un sujeto u objeto en un conjunto de variables observadas.
K es el número de clases.
π_k indica la probabilidad de pertenecer a una clase latente k (tamaño de la clase k).
J indica el número total de indicadores.
j un indicador particular.
$f_k\ (y_{ij} \mid \theta_{jk})$ es la función de distribución univariante de cada uno de los elementos y_{ij} de y_i, condicionada por el conjunto de variables indicadoras j de la clase k.[19]

Los parámetros del modelo de clases latentes se estiman por el método de máxima verosimilitud, es decir, la solución consiste en valores de parámetros que maximizan la función de probabilidad y su logaritmo natural (Uebersax, 2000). La verosimilitud de un modelo se define como la probabilidad de que cada conjunto de datos haya sido generado por el modelo; es la formulación del modelo a través de la distribución conjunta de los datos y se expresa de la siguiente forma:

$$Ln\ (\theta) = \prod_{i=1}^{n} f\ (y_i \mid \theta) = \sum_{i=1}^{n} log\, f\ (y_i \mid \theta)$$

Donde:

y_i representa un conjunto de datos particular,
n es el número de casos θ y comprende los parámetros del modelo.

El análisis de clases latentes en cuanto a objetivos es semejante al análisis de clúster, su diferencia radica en que el primero asigna los sujetos a las clases, basándose en una función de la probabilidad de pertenencia y el de clúster los asigna a base de distancias. Además, el análisis de clases latentes hace posible la elección del número de clases de forma menos arbitraria, no requiere estandarización, y, al igual que el

[19] Es decir, la función de densidad de un conjunto de respuestas de un sujeto en un conjunto de variables observadas es igual a la suma de la probabilidad de pertenecer a cada una de las clases por el producto de la función de densidad de cada indicador condicionado por la clase.

análisis factorial, resulta útil para la reducción de datos. Determinar el número de clases es semejante a determinar el número de factores; su objetivo es hallar un equilibrio entre el ajuste a los datos y el número de clases o factores requerido (Uebersax, 2009).

Resultados a partir de la aplicación de las metodologías propuestas

En este artículo se presentan algunos resultados de los métodos sugeridos, enfocados solo al cálculo del índice sin considerar la estratificación. La existencia de varias propuestas demuestra el interés que existe al respecto. A continuación, se analizan algunas metodologías que reúnen ciertas características, entre ellas:

- Sencillez en su estimación,
- Comparabilidad en el tiempo,
- Comparabilidad entre unidades territoriales, y
- Solidez metodológica y estadística.[20]

Media aritmética (IAM)

La aplicación de esta técnica permite comparar en el tiempo la evolución de la variabilidad entre los indicadores, los cambios y la posición entre unidades. En la gráfica 1 se observa la posición resultante en la aplicación tanto del índice de marginación actual, como del índice absoluto de marginación, mediante la estimación de una media aritmética.

La diferencia entre las posiciones de las entidades federativas en el índice de marginación y el índice absoluto de marginación (IAM) muestra cambios en 20; once de ellas ganan posiciones y mejoran su *ranking* en la marginación, como Baja California Sur que pasa del lugar 23, empleando el índice de marginación, al sitio 28 con el IAM; o bien Aguascalientes, que desciende del lugar 28 al 24, siendo una de las nueve entidades federativas que pierden posición; doce entidades se mantuvieron iguales (véase cuadro 4). La correlación entre ambos índices es muy cercana a la unidad (0.99) y el rango estadístico de los datos[21] va de 4 en el caso del índice actual a 23.8 en el IAM.

Media geométrica

Para este caso, al seguir la propuesta del IDH, se contempló la creación de cuatro componentes, generados para cada dimensión empleada en el cálculo del índice de marginación: educación, vivienda, distribución de la población e ingresos monetarios. Dado que los indicadores son calculados como porcentajes; los valores máximos de cada indicador son cien y los mínimos cero, considerando la formulación de cada componente en el IDH:

$$\textit{Índice del componente} = \frac{\textit{valor real - mínimo}}{\textit{máximo - mínimo}}$$

Si el cálculo fuera directo, los componentes solo cambiarían en términos de unidad, por lo que convendría establecer cotas superiores e inferiores respecto a cada dimensión, analizando todas las unidades; aunque este valor serviría para su comparación en el tiempo, se necesita respaldar estos criterios determinados por instancias internacionales. Para este artículo se reserva su estimación por no contar con criterios para determinar dichas cotas.

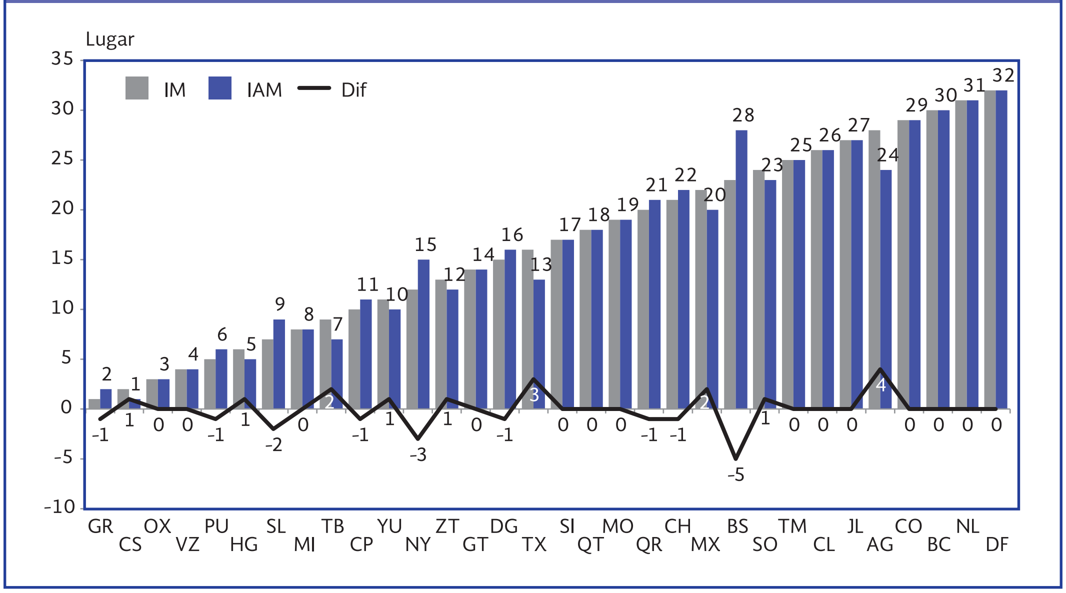
Distancias euclidianas (índice de marginación euclidiano)

Al aplicar esta propuesta y contrastarla con la del índice de marginación actual, solo siete entidades mantuvieron su posición (véase gráfica 2), el resto muestra grandes cambios, pues 13 entidades federativas ganan

[20] Aunque no es adecuada la comparación directa entre un ejercicio y otro por el inherente carácter definido por cada metodología, el sentido de mostrar la posición que ocupa cada unidad geográfica es meramente ilustrativo, no marcando la intensidad o algún indicativo de mejoría entre aplicaciones.

[21] Rango estadístico (R) o recorrido estadístico es el intervalo entre el valor máximo y el valor mínimo; por ello, comparte unidades con los datos. Permite obtener una idea de la dispersión de los datos, cuanto mayor es el rango, más dispersos están los datos de un conjunto.

Gráfica 1.
Posición que obtienen las entidades federativas tanto en el índice actual como al emplear la media aritmética y su diferencia, 2010



Fuente: Estimaciones del CONAPO con base en el INEGI, Censo de Población y Vivienda 2010.

lugar, como Baja California Sur que pasó del sitio 23 en el índice actual al 26 en el IME, o como Aguascalientes y Tlaxcala que de la posición 28 y 16 se movieron a la 22 y 10, respetivamente; en tanto, doce entidades pierden posición respecto al índice actual.

En el cuadro 4 se advierte que las variaciones existentes entre el método propuesto y el actual son considerables; vale la pena señalar que la correlación existente entre los dos índices es de 0.965, lo que indica el parecido con el índice actual y se tiene un rango de 79.8 puntos, lo que señala la dispersión resultante.

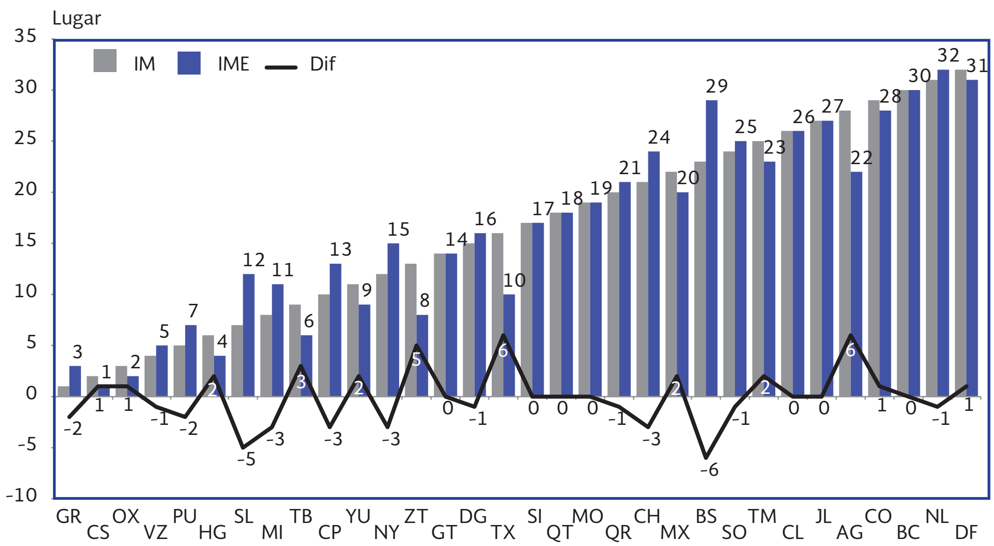
Componentes principales mediante la matriz de covarianza

En el ejercicio de Análisis de Componentes Principales que emplea la matriz de covarianzas, las varianzas de cada una de las variables de estudio se encuentran en la diagonal de dicha matriz, y las covarianzas, en la matriz triangular superior o inferior, tomando en cuenta que es una matriz simétrica (véase cuadro 1). Dado que la matriz analizada es la de varianzas-covarianzas, se consideró el número de veces que un autovalor es mayor que 1[22] de la matriz para que el factor sea considerado en la solución.

Al analizar esta propuesta, se distingue que la proporción de varianza explicada por la primera componente es de 68.2 por ciento. Con la aplicación del método de componentes principales mediante la matriz de covarianzas, la influencia de las ponderaciones cambia de manera importante. En el cuadro 2 se puede apreciar que los indicadores de población en localidades con menos de cinco mil habitantes (0.96), población de 15 años o más sin primaria completa (0.92), población ocupada con ingreso de hasta dos salarios mínimos (0.92), población de 15 años o más analfabeta (0.91), ocupantes en viviendas sin agua entubada (0.82) y ocupantes en viviendas con piso de tierra (0.81) resultan significativos, mientras que con menor representatividad se encuentran los porcentajes de: viviendas con algún nivel de hacinamiento, ocupantes en viviendas sin energía eléctrica y ocupantes en viviendas sin drenaje ni excusado.

[22] El autovalor promedio, por defecto es 1, pero se puede cambiar entre cero y el número de variables.

Gráfica 2.
Posición que obtienen las entidades federativas tanto en el índice actual como al emplear distancias euclidianas y su diferencia, 2010



Fuente: Estimaciones del CONAPO con base en el INEGI, Censo de Población y Vivienda 2010.

Debido a que el análisis de componentes principales que utiliza la matriz de covarianza devuelve solamente un componente con autovalor mayor que 1, al ordenar las unidades y compararlas con el índice actual, solo se registran coincidencias en nueve entidades federativas. En la gráfica 3 se pueden identificar los cambios al implementar la matriz de covarianzas, donde el panorama difiere del observado por el índice de marginación actual; en 23 entidades se presentan posiciones distintas, de nuevo Aguascalientes y Baja California Sur quedan en los extremos, como las entidades que más pierden o ganan posiciones con respecto al índice actual.

En el cuadro 4 se exponen los resultados obtenidos en este ejercicio, donde el rango de variación de los datos es de 3.8 puntos, menor al del índice actual (4.0). La implementación de esta propuesta arroja una correlación de 0.961 entre los índices, lo que permite asegurar lo relacionados que se encuentran.

Análisis Factorial Confirmatorio

La aplicación del Análisis Factorial Confirmatorio se realizó en dos etapas, empleando los paquetes estadísticos SPSS y EQS (Structural Equation Modeling Software). Primero, se efectuó el análisis factorial exploratorio y, posteriormente, el cálculo de la invarianza factorial; con ambos es posible asegurar que la aplicación del modelo resulte favorable.

La aplicación de esta técnica determinó que con un solo factor se resume el 70.3 por ciento de la varianza total explicada, con un coeficiente Kaiser-Meyer-Olkin (KMO) de 0.89 que para fines del proceso se manifiesta favorable. Por lo anterior, es posible asegurar que su implementación es conveniente.

Al utilizar el método de máxima verosimilitud para estimar las cargas factoriales, la matriz factorial obtenida muestra cargas superiores a 0.5 en cada uno de los indicadores; para continuar con la implementación del método se consideraron los nueve indicadores

Cuadro 1.
Matriz de Covarianza

Indicadores	% Población de 15 años o más analfabeta	% Población de 15 años o más sin primaria completa	% Ocupantes en viviendas sin drenaje ni excusado	% Ocupantes en viviendas sin energía eléctrica	% Ocupantes en viviendas sin agua entubada	% Viviendas con algún nivel de hacina-miento	% Ocupantes en viviendas con piso de tierra	% Población en localidades con menos de 5 000 habitantes	% Población ocupada con ingreso de hasta 2 SM
% Población de 15 años o más analfabeta	17.44								
% Población de 15 años o más sin primaria completa	26.42	45.05							
% Ocupantes en viviendas sin dre-naje ni excusado	9.63	14.7	14.59						
% Ocupantes en viviendas sin energía eléctrica	3.42	5.6	2.47	1.76					
% Ocupantes en viviendas sin agua entubada	25.88	38.04	13.68	6.47	52.02				
% Viviendas con algún nivel de hacinamiento	22.8	33.29	13.98	3.94	34.73	44.9			
% Ocupantes en viviendas con piso de tierra	17.34	26.02	8.15	4.26	30.82	21.46	21.94		
% Población en localidades con menos de 5 000 habitantes	53.94	90.67	31.29	12.7	84.95	70.58	53.36	262.8	
% Población ocu-pada con ingreso de hasta 2 SM	41.58	67.15	24.34	8.54	56.47	61.33	37.53	153.33	136.54

Fuente: Estimaciones del CONAPO con base en el INEGI, Censo de Población y Vivienda 2010.

que resultaron significativos.[23] En el cuadro 3 se exponen las ecuaciones resultantes del modelo.

Al ordenar y comparar los indicadores, únicamente diez entidades federativas conservaron su sitio, el resto cambió de lugar; destacan Durango y Colima que fueron las que más posiciones avanzaron o retrocedieron, respectivamente (véase gráfica 4). Así, la aplicación del Análisis Factorial Confirmatorio permite apreciar cambios diferentes a los que se han observado en el resto de las metodologías propuestas, teniendo una alta correlación (0.98) con el índice actual.

Dichas estimaciones permiten observar que, dada la alta relación entre estos índices y el rango de 3.8, es similar al obtenido en el índice actual, la aplicación de este método puede constituir una opción viable, aunque ciertamente habrá que considerar la necesidad de contar con conocimientos de estadística y matemáticas. Los datos se pueden consultar en el cuadro 4.

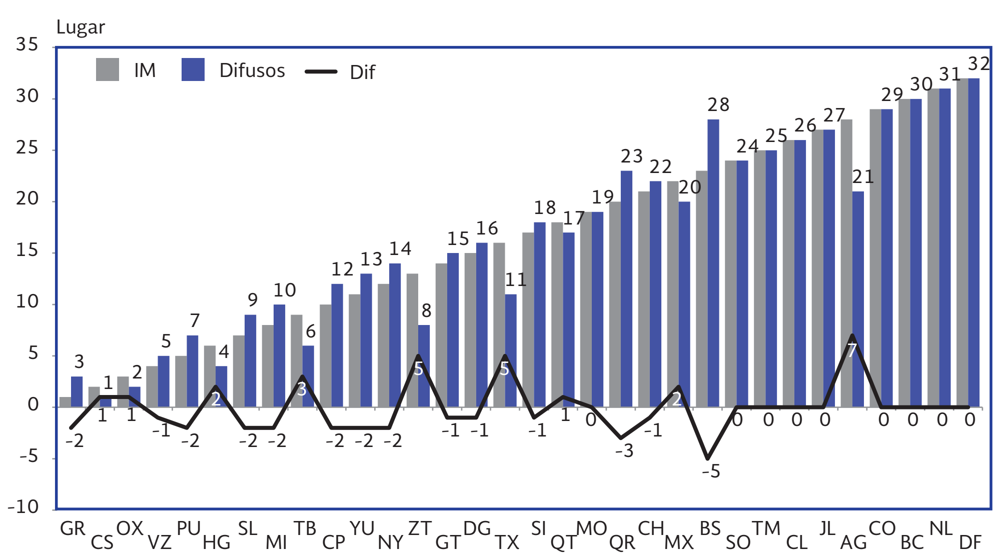
23 En el Análisis Factorial Confirmatorio conviene considerar todas aquellas cargas superiores a 0.7, pues los resultados serán mejores.

Cuadro 2.
Matriz del componente reescalado

Indicadores	Componente reescalado
% Población en localidades con menos de 5 000 habitantes	0.96
% de Población de 15 años o más sin primaria completa	0.92
% Población ocupada con ingreso de hasta 2 salarios mínimos	0.92
% de Población de 15 años o más analfabeta	0.91
% Ocupantes en viviendas sin agua entubada	0.82
% Ocupantes en viviendas con piso de tierra	0.81
% Viviendas con algún nivel de hacinamiento	0.79
% Ocupantes en viviendas sin energía eléctrica	0.64
% Ocupantes en viviendas sin drenaje ni excusado	0.58

Fuente: Estimaciones del CONAPO con base en el INEGI, Censo de Población y Vivienda 2010.

Gráfica 3.
Posición que obtienen las entidades federativas tanto en el índice actual como del análisis de componentes principales al emplear la matriz de covarianzas y su diferencia, 2010



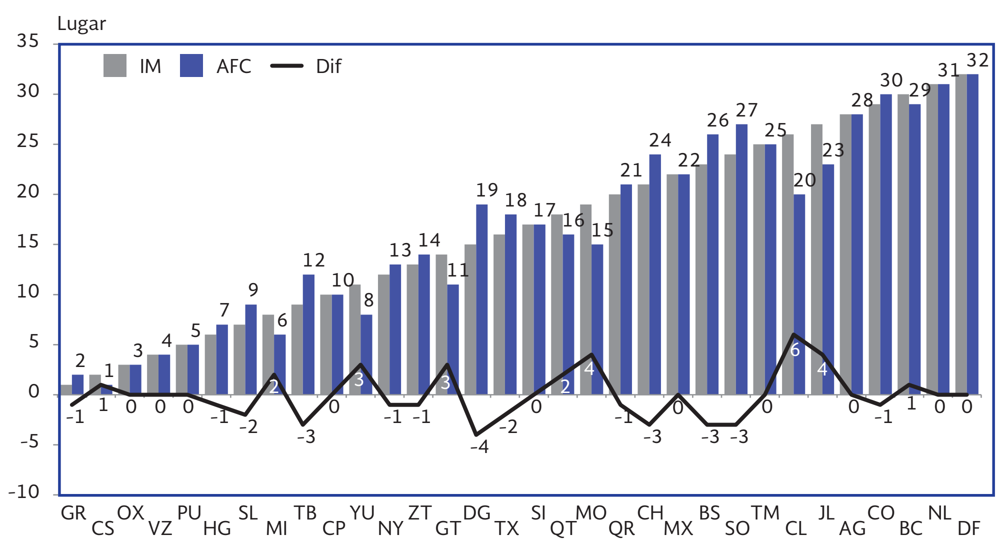
Fuente: Estimaciones del CONAPO con base en el INEGI, Censo de Población y Vivienda 2010.

Cuadro 3.
Ecuaciones y varianza explicada por cada indicador del modelo

Ecuaciones del modelo con la contribución de cada indicador	Varianza explicada del factor por cada indicador
IND1 = V4 = 0.987*F1 + 0.161 E4	0.974
IND2 = V5 = 0.952*F1 + 0.305 E5	0.907
IND3 = V6 = 0.606*F1 + 0.795 E6	0.367
IND4 = V7 = 0.651*F1 + 0.759 E7	0.423
IND5 = V8 = 0.871*F1 + 0.492 E8	0.758
IND6 = V9 = 0.816*F1 + 0.579 E9	0.665
IND7= V10 = 0.893*F1 + 0.450 E10	0.797
IND8 = V11 = 0.826*F1 + 0.564 E11	0.682
IND9 = V12 = 0.866*F1 + 0.501 E12	0.749

Fuente: Estimaciones del CONAPO con base en el INEGI, Censo de Población y Vivienda 2010.

Gráfica 4.
Posición que obtienen las entidades federativas tanto en el índice actual como al emplear el análisis factorial confirmatorio y su diferencia, 2010



Fuente: Estimaciones del CONAPO con base en el INEGI, Censo de Población y Vivienda 2010.

Cuadro 4.
Comparación entre el valor del índice actual y el de los métodos propuestos, 2010

Entidad federativa	Índice de margi-nación	Lugar	Índice absoluto	Lugar	Distancia euclidiana	Lugar	ACP con matriz de covarian-zas	Lugar	Análisis factorial confirma-torio	Lugar	Con-juntos difusos	Lugar
Aguascalientes	-0.91	28	12.40	24	54.03	22	-0.59	21	-0.85	28	0.04	28
Baja California	-1.14	29	9.47	30	40.38	30	-1.30	30	-1.04	29	0.04	29
Baja California Sur	-0.68	23	11.65	28	45.84	29	-0.99	28	-0.80	26	0.05	23
Campeche	0.43	10	19.61	11	76.67	13	0.29	12	0.35	10	0.09	10
Coahuila	-1.14	29	10.19	29	46.12	28	-1.11	29	-1.06	30	0.03	30
Colima	-0.78	26	12.07	26	51.08	26	-0.86	26	-0.49	20	0.05	26
Chiapas	2.32	2	31.51	1	116.69	1	2.22	1	2.53	1	0.15	3
Chihuahua	-0.52	21	12.90	22	52.13	24	-0.72	22	-0.68	24	0.05	21
Distrito Federal	-1.48	32	7.68	32	39.72	31	-1.59	32	-1.28	32	0.02	32
Durango	0.05	15	17.20	16	67.25	16	0.10	16	-0.43	19	0.08	14
Guanajuato	0.06	14	17.77	14	69.21	14	0.14	15	0.24	11	0.08	15
Guerrero	2.53	1	30.73	2	104.70	3	1.71	3	2.32	2	0.18	1
Hidalgo	0.66	6	22.61	5	89.97	4	1.18	4	0.70	7	0.11	5
Jalisco	-0.82	27	11.83	27	48.18	27	-0.87	27	-0.62	23	0.05	25
México	-0.55	22	13.85	20	57.75	20	-0.58	20	-0.59	22	0.05	22
Michoacán	0.53	8	20.49	8	77.61	11	0.56	10	0.79	6	0.10	9
Morelos	-0.27	19	15.58	19	61.09	19	-0.27	19	-0.14	15	0.07	18
Nayarit	0.12	12	17.75	15	68.83	15	0.20	14	-0.04	13	0.08	12
Nuevo León	-1.38	31	7.97	31	36.89	32	-1.57	31	-1.21	31	0.03	31
Oaxaca	2.15	3	29.78	3	108.05	2	1.98	2	2.23	3	0.16	2
Puebla	0.71	5	22.01	6	85.00	7	0.77	7	0.82	5	0.10	8
Querétaro	-0.26	17	15.81	18	62.64	18	-0.07	17	-0.24	16	0.07	16
Quintana Roo	-0.42	20	13.59	21	57.05	21	-0.77	23	-0.49	21	0.06	20
San Luis Potosí	0.56	7	20.39	9	76.74	12	0.57	9	0.38	9	0.10	7
Sinaloa	-0.26	17	15.91	17	63.57	17	-0.14	18	-0.33	17	0.07	17
Sonora	-0.70	24	12.44	23	52.08	25	-0.80	24	-0.82	27	0.05	24
Tabasco	0.47	9	21.84	7	86.08	6	0.97	6	0.24	12	0.10	6
Tamaulipas	-0.72	25	12.35	25	53.63	23	-0.83	25	-0.72	25	0.04	27
Tlaxcala	-0.15	16	18.00	13	79.19	10	0.36	11	-0.33	18	0.06	19
Veracruz	1.08	4	23.84	4	88.38	5	1.08	5	1.15	4	0.12	4
Yucatán	0.42	11	19.62	10	79.24	9	0.25	13	0.49	8	0.09	11
Zacatecas	0.10	13	19.60	12	80.39	8	0.70	8	-0.07	14	0.08	13

Fuente: Estimaciones del CONAPO con base en el INEGI, Censo de Población y Vivienda 2010.

Análisis multivariado mediante conjuntos difusos

Este método tiene una muy alta correlación, de 0.992, con el índice actual. Su elaboración es de fácil aplicación ya que únicamente requiere de conocimientos matemáticos básicos. Según se ilustra, conserva 17 entidades federativas en la misma posición y el resto varía máximo en tres lugares; estos movimientos los presentan Puebla, que pasa de la posición 5 a la 8, y Tabasco, de la 9 a la 6 (véase gráfica 5).

En el cuadro 4 se exhiben los valores de este indicador, los cuales evidencian un rango de variación de 0.16, mostrando la alta concentración de los datos.

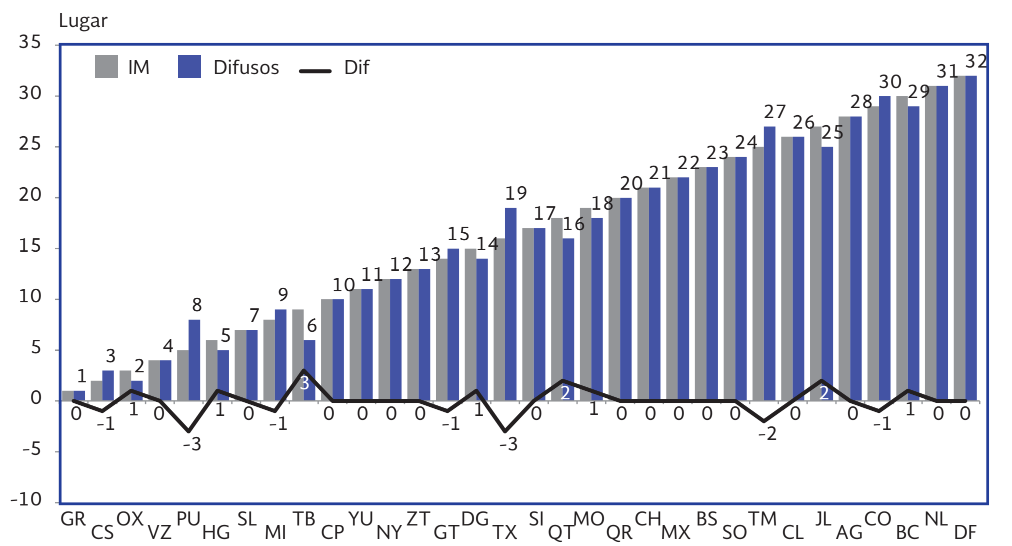
Consideraciones finales

La relevancia que ha tenido el índice de marginación como fuente para la determinación de un importante número de acciones en materia de política social se ve reflejada en los poco más de 300 programas o proyectos enfocados a la distribución de recursos y aplicación de estrategias que lo utilizan. Así, el análisis de las distintas alternativas metodológicas responde al compromiso del CONAPO con la mejora de los insumos sociodemográficos que genera. En este sentido, el cálculo de la marginación constituye una de las grandes tareas de la institución, por su innegable vinculación con la política de población y su impacto en la instrumentación de políticas públicas que derivan en apoyos para aquellos mexicanos que habitan en unidades geográficas clasificadas en los grados más altos de marginación, de ahí la relevancia de tener en perspectiva las posibles mejoras metodológicas, sin perder de vista la solidez conceptual y metodológica.

Aunque en un principio el índice se vio respaldado por la metodología innovadora, es claro que en la actualidad se ve parcialmente rebasado por los requerimientos de algunos actores y acciones de política pública que lo utilizan, especialmente en cuanto a la comparabilidad en el tiempo, sin dejar de lado la pertinencia de los indicadores empleados y el impacto político que tendría un cambio metodológico significativo.

Cabe mencionar que el análisis de las diversas propuestas metodológicas surgió del acercamiento

Gráfica 5.
Posición que obtienen las entidades federativas tanto en el índice actual como al emplear conjuntos difusos y su diferencia, 2010



Fuente: Estimaciones del CONAPO con base en el INEGI, Censo de Población y Vivienda 2010.

Cuadro 5.
Comparación entre las características esperadas para la adopción de una nueva metodología

Características	ACP con matriz de correlación	Media aritmética	Media geomé-trica	Distancias euclidianas	ACP con matriz de covarianza	Análisis factorial confirma-torio	Conjuntos difusos	Análisis de clases latentes
Reduce la dimensión del fenómeno.	X	X	X	X		X	X	X
Establece un orden entre las unidades geográficas.	X	X	X	X	X	X	X	X
Permite la comparación en el tiempo para evaluar la marginación.		X	X	X		X	X	X
Comparable entre unidades de distinto nivel geográfico con indicadores similares.		X	X	X		X	X	X
No requiere de un software especializado para realizar los cálculos.		X	X	X			X	
De fácil interpretación.		X					X	
Los datos extremos no influyen en su valor.	X		X	X	X	X	X	X
No requiere conocimientos especializados.		X	X				X	
Los ponderadores señalan la afectación en los indicadores.	X				X	X	X	X
Es posible calcularlo para los cuatro niveles geográ-ficos (Entidad federativa, municipio, localidad y AGEB urbana).	X	X	X	X	X	X	X	
Total	**5**	**8**	**8**	**7**	**4**	**7**	**10**	**6**

Fuente: Estimaciones del CONAPO.

del CONAPO con especialistas, tanto del sector público como del académico, para recibir retroalimentación. Los resultados presentados proporcionan un horizonte similar al de la aplicación actual, lo que es un evidente indicio de que una mejora no implica necesariamente cambios profundos en los mismos; aunque la institución está consciente de la necesidad de realizar adecuaciones a este índice.

Otro hallazgo del trabajo es que si bien los métodos analizados implican ventajas y desventajas, tal como sucede con la forma actual de cálculo, éstos abordan por primera vez la posibilidad de comparar en el tiempo la presencia de la marginación en las unidades territoriales de estudio. A manera de resumen, el cuadro 5 detalla algunas de las características consideradas en el análisis de las metodologías alternativas al índice de marginación calculado mediante el método de Análisis de Componentes Principales empleando la matriz de correlaciones. Dados los rubros contemplados, no es de extrañar que la forma actual de cálculo sea una de las que obtiene menor puntuación. En contraparte, de acuerdo a esta valoración, el cálculo que resulta más conveniente es el de la aplicación de conjuntos difusos.

Por último, es necesario señalar que también es de suma importancia la revisión de la parte conceptual y de la estratificación, esta última, por ejemplo, es fundamental dado el empleo del grado en la aplicación de políticas públicas. Una metodología simple en su aplicación pero sólida en sus resultados establecerá una clara mejora en la medición del fenómeno.

Bibliografía

Bollen, Kenneth A. (1989), *Structural Equations with Latent Variables*, New York, Willey.

——— y Curran P. J. (2006), *Latent Curve Models. A Structural Equation Perspective*, Willey, pp. 248-252.

Bustos, Alfredo (2009), "Niveles de marginación: Una propuesta técnica". Ponencia presentada en el Seminario de Actualización del Marco Conceptual y Metodológico del Índice de Marginación, junio 2012, no publicada, 25 pp.

——— (2011), "Niveles de marginación: una estrategia multivariada de clasificación. Realidad, datos y espacio" en INEGI, *Revista Internacional de Estadística y Geografía*, vol. 2, núm. 1

CONAPO (2004), *Índice absoluto de marginación 1990-2000*, Consejo Nacional de Población, México. Colección: índices sociodemográficos, México. Disponible en línea: http://www.conapo.gob.mx/es/CONAPO/Indice_ absoluto_de_ marginacion_1990-2000_

COPLAMAR, Coordinación General del Plan Nacional de Zonas Deprimidas y Grupos (1977), *Bases para la acción, 1977-1982*, mimeo, Palacio Nacional, México, 1977, 10 pp.

Dalenius, T. y J. Hodges (1959), "Minimum Variance Stratification", en *Journal of the American Statistical Association*, vol. 54, núm. 285, pp. 88-101.

González Rosas, Javier y José Luis Torres Islas (2011), "Evaluación de la marginación en México 2000-2005", documento interno no publicado, 7 pp.

Goodman, L. A. (1974), *Exploratory latent structure analysis using both identifiable and unidentifiable models. Biometrika*, 1974, pp. 61:215-31.

Harman, Harry H. (1976), *Modern Factor Analysis*, Chicago, University of Chicago Press, 3ª ed., 1976.

McArdle, John J. (1994), "Structural factor analysis experiments with incomplete data" en *Multivariate Behavioral Research*, vol.29, núm.4, pp. 409-454.

McArdle, John J. & D. Epstein (1987), "Latent growth curves within developmental structural equation models" en *Child Development*, vol. 58, núm. 1, pp. 110-133.

Morales-Ramos, Marco Antonio y Eduardo Morales-Ramos (2008), "La teoría de conjuntos difusos como una opción para medir la pobreza", en *El Trimestre Económico*, vol. LXXV (3), núm. 299, pp. 641-662.

Mulaik, S. A. (2010), *Foundations of Factor Analysis*, Second Edition, Chapman & Hall/CRC, Boca.

Neter, J. y W. Wasserman (1974), *Applied Linear Statistical Models*, Richard D. Irwin Inc., Homewood, Illinois.

PNUD (2003), *Informe sobre Desarrollo Humano México 2002*, PNUD, México.

Uebersax J. S. (2009), *Latent class analysis frequently asked questions* (FAQ). Disponible en línea: http://www.john-uebersax.com/stat/faq.htm.

——— (2000), A brief study of local maximum solutions in latent class analysis. Disponible en línea: http://john-uebersax.com/stat/local.htm