

Construcción de un índice compuesto y aproximación para medir los cambios en el tiempo

José Vences Rivera

En este trabajo se desarrolló una metodología para la construcción de un índice compuesto que resume la información contenida en fenómenos de naturaleza multidimensional. La varianza explicada por este índice es casi igual a la obtenida por la técnica de análisis de componentes principales (ACP) respecto a la varianza total del conjunto original de variables consideradas, lo cual fue corroborado mediante simulación de Monte Carlo y con aplicación a datos derivados de los censos de población y vivienda levantados por el Instituto Nacional de Estadística y Geografía (INEGI). El resultado de los cambios en el tiempo se obtiene al establecer un espacio común de mediciones.

Las ventajas del nuevo índice comparado con el de la primera componente principal son: el cálculo es muy simple e intuitivo; es de fácil interpretación, ya que los resultados se dan en la misma escala que las variables de insumo; y no se prevén problemas de comunicación con el usuario en general.

Palabras clave: índice multivariado, marginación, primera componente principal, varianza explicada, ponderador, proyección ortogonal.

Recibido: 13 de noviembre de 2013

Aceptado: 24 de abril de 2014

This paper shows the development of a methodology aimed for the creation of a compound index that summarizes the information contained in multidimensional phenomena. The variance explained according to this index is almost identical to that obtained with the technique of Principal Components Analysis (PCA) in regard to the total variance of the original set of considered variables. This was proven by means of a Monte Carlo simulation as well as by applying data that derived from Population and Housing Censuses conducted by the National Institute of Statistic and Geography (INEGI). The result of changes over time is obtained by having established a common space of measurements.

The advantages of the new index, compared with that of the first main component, are that the calculation is very simple and intuitive; it is of easy interpretation since the results are given in the same scale that the input variables; and that we do not foresee communication problems with the user.

Key words: multivariate index, marginalization, first main component, explained variance, weight, orthogonal projection.



Rafting. Siberian River. Kazur. Kuznetsky Alatau ©iStockphoto.com/isoft

1. Introducción

Los indicadores sintéticos son de gran utilidad para resumir la información contenida en un conjunto, generalmente grande, de variables medidas sobre un mismo individuo u objeto; permiten brindar un primer referente de la situación que prevalece en fenómenos de naturaleza multidimensional y, así, orientar a quienes toman decisiones en los diversos campos de acción. Un indicador sintético debe ser fácil de calcular, confiable, comparable en el tiempo y en el espacio, sencillo de interpretar y de fácil comunicación con el usuario en general; para su construcción, suelen utilizarse técnicas de análisis estadístico multivariado para analizar, de manera simultánea, tres o más variables. Entre las más usuales se ubica el análisis de componentes principales (ACP) en sus diferentes variantes.

En la revisión de literatura especializada (al menos en el campo de la estadística oficial), no se encontró ningún procedimiento para construir un índice multivariado que cumpla con las características mencionadas. En particular, se analizaron las técnicas de factores dinámicos, STATIS y componentes principales comunes (ver Flury, 1984 y 1987; Forni, 2000; Lavit, 1982 y 1988; Watson, 1983; Geyne y Hernández, 2007).

Éstas son de las más citadas para establecer comparaciones en el tiempo, y los resultados obtenidos mediante su aplicación son válidos e inamovibles sólo en el periodo considerado; sin embargo, el problema surge cuando se quiere incorporar información de otro tiempo; en este caso debe considerarse el nuevo conjunto de datos históricos agrandado, y los resultados cambian respecto a los obtenidos en

el cálculo anterior, lo cual puede generar desconfianza y desconcierto tanto en los usuarios como en los tomadores de decisiones, quienes podrían estar familiarizados con las cifras proporcionadas en algún ejercicio previo. La explicación de esta situación al usuario no es fácil de comunicar, además de que las técnicas en sí tienen un alto grado de especialización.

Cortés y Vargas (2011) proponen una metodología para medir la marginación en México a través del tiempo mediante el uso del análisis factorial confirmatorio y del concepto de invarianza longitudinal. De la lectura de su artículo se desprende que los puntajes obtenidos no están acotados entre cero y cien (en términos porcentuales) como lo están los indicadores básicos que sirven de insumo para la construcción del índice compuesto; su cálculo no es simple, se requiere cierto grado de adiestramiento, pero tal vez la mayor área de oportunidad que se observa es la difícil comunicación con el usuario.

La Organización para la Cooperación y el Desarrollo Económicos (OCDE, 2008) desarrolló un manual a manera de guía para la construcción y uso de indicadores compuestos; constituye una amplia revisión de aspectos conceptuales y prácticos, y proporciona recomendaciones que ayudan a entender sus aplicaciones y la interpretación de resultados.

Los índices multivariados construidos en México han sido un referente importante para el diseño y evaluación de programas sociales, así como para la focalización de recursos; permiten identificar, particularmente, a las unidades geográficas con mayores carencias y dan la pauta para realizar estudios específicos sobre los grupos de población vulnerable que son objeto de apoyos gubernamentales. A partir de 1990 se identifican los siguientes índices:

- Índice de rezago social para reubicar las tiendas del Sistema de Distribuidoras Conasupo, SA de CV (DICONSA).

- Índice para medir la calidad del empleo en las ciudades de México. Universidad Autónoma del Estado de Hidalgo (UAEH).
- Índice para medir el cumplimiento de los derechos humanos del Instituto Nacional de Derechos Humanos (INDH).
- Índice de marginación del Consejo Nacional de Población (CONAPO).
- Índice de rezago social del Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL).
- Índice de rezago del Gobierno del Distrito Federal.

Lo anterior originó el desarrollo del presente trabajo, que tiene como objetivo construir una medida resumen sencilla que dé cuenta de la magnitud del fenómeno estudiado y de sus cambios en el tiempo, que facilite la interpretación de resultados y la comunicación con el usuario, para apoyar las decisiones dirigidas a mejorar el bienestar de la población.

2.- Marco teórico

En el ACP existen tantas componentes principales (CP) independientes como variables de insumo correlacionadas se hayan considerado. La primera CP explica la mayor cantidad posible de la varianza conjunta de esas variables para una combinación lineal, y expresa el *tamaño promedio* del fenómeno estudiado, razón por la cual suele utilizarse como un índice cuyo nombre es acorde con la naturaleza del problema, y la calidad del mismo se fortalece cuanto mayor es la varianza explicada; éste se construye al optimizar una función de varianzas y es claro que ningún otro que sea combinación lineal de variables puede contener más información que la primera CP.

Por lo anterior, la técnica de ACP ha sido de las más utilizadas para la construcción de índices socioeconómicos. No obstante, la relevancia teórica del índice resultante presenta en la práctica ciertas limitaciones; si bien es cierto que retiene la máxi-

ma cantidad de información —es decir, es el que mejor *representa* al conjunto original de variables y permite un ordenamiento entre las unidades de observación—, no es de utilidad para medir los cambios en el tiempo y tampoco es fácil de interpretarlo ni comunicar los resultados con el usuario en general.

Así, para cumplir con la primera parte del objetivo planteado, se ha generado un nuevo índice que conserva, en la medida de lo posible, las propiedades teóricas de la primera CP y supera las limitaciones mencionadas. Para ello, la metodología utilizada en este trabajo fue la siguiente:

- Se estudió a profundidad la técnica de análisis de componentes principales aplicada a la matriz de correlaciones R (Jackson, 1991; Johnson, 1982; Peña, 2003; Vences, 1999). Algo nuevo desprendido de ello es que cuando las variables de insumo forman un solo factor, en la combinación lineal de la primera CP el mayor ponderador corresponde a la variable que, en promedio, está más correlacionada con el resto; el segundo ponderador, en orden decreciente de magnitud, es el asociado con la variable que presenta la segunda mayor correlación con las demás; y así, de forma sucesiva. Esto es, el vector característico de la primera CP es aproximadamente proporcional al vector de correlaciones promedio entre cada variable original y el resto.
- Se generaron varios índices alternativos.
- Los diversos índices se compararon con la primera CP (Y_1) en términos de la varianza explicada. Para esto, en principio se recurrió a la técnica de simulación de Monte Carlo; después, se realizaron aplicaciones con datos derivados de los censos de población y vivienda 1990-2010 (INEGI, 2012) y se contrastaron con los resultados esperados.
- El mejor índice sería aquel que en términos de la información retenida fuera similar al obtenido por el ACP, pero de fácil interpretación y simplicidad en su cálculo y, sobre todo, que resultara de utilidad para el usuario.

Por su parte, para la medición aproximada de los cambios en el tiempo se utilizó la proyección ortogonal de un vector sobre otro y, sin pérdida de generalidad, para referirse al tiempo se ha usado el término *año*.

3. El índice

Sea p = número de variables correlacionadas: x_1, x_2, \dots, x_p .

El índice es una combinación lineal de la forma:

$$I = c_1x_1 + c_2x_2 + \dots + c_px_p$$

donde,

$$C_i = \frac{r_i}{s_i s}, \quad S_i = \text{desviación estándar de la variable } i,$$

$$s = \frac{r_1}{s_1} + \frac{r_2}{s_2} + \dots + \frac{r_p}{s_p},$$

y r_i = media cuadrática de las correlaciones entre la variable i y el resto de las variables, dada por:

$$r_i = \left(\frac{1}{p-1} \sum_{j=1}^p r_{ij}^2 \right)^{1/2}$$

esto para $i = 1, 2, \dots, p$; con $i \neq j$.

Así, el índice toma la forma:

$$I = \left(\frac{r_1}{s_1 s} \right) x_1 + \left(\frac{r_2}{s_2 s} \right) x_2 + \dots + \left(\frac{r_p}{s_p s} \right) x_p = \sum_{i=1}^p \frac{r_i}{s_i s} x_i$$

o bien,

$$I = \frac{\left(\frac{r_1}{s_1} \right) x_1 + \left(\frac{r_2}{s_2} \right) x_2 + \dots + \left(\frac{r_p}{s_p} \right) x_p}{\frac{r_1}{s_1} + \frac{r_2}{s_2} + \dots + \frac{r_p}{s_p}}$$

El índice constituye un promedio aritmético ponderado, donde la suma de los ponderadores es igual a la unidad. Por lo tanto, se interpreta en términos de las unidades de las variables originales y

es muy fácil de calcular por el usuario en general; aspectos de los que adolece la técnica de ACP.

4. Los cambios en el tiempo

En la realidad, el caso más usual es que la matriz de correlaciones (o de covarianzas) entre las variables consideradas no permanezca constante en el transcurso del tiempo debido a que los cambios registrados en las unidades de observación no son proporcionales, de manera que los vectores de ponderaciones resultantes son diferentes y, por lo tanto, apuntan en direcciones también distintas, lo cual hace que sus magnitudes no sean directamente comparables. Para salvar esta situación, una vez que el nuevo índice es obtenido para cada año, se calcula la norma de la proyección ortogonal a un año base, con el fin de tener un espacio común de mediciones de una sola dimensión y, así, determinar los cambios aproximados en el tiempo. La proyección se logra mediante el producto escalar:

$$\text{Proy}_u(v) = \frac{u \cdot v}{u \cdot u} u$$

y la norma o magnitud de este vector se calcula como:

$$\frac{|u \cdot v|}{\|u\|}$$

donde,

u = vector de ponderaciones en la combinación lineal del nuevo índice para el año base definido.

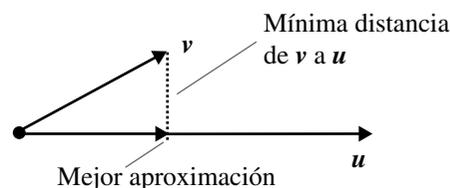
v = vector de ponderaciones en la combinación lineal del nuevo índice para los otros años.

$\text{Proy}_u(v)$ = proyección ortogonal del vector v sobre el vector u .

$|u \cdot v|$ = valor absoluto del producto escalar entre los vectores, u y v .

$\|u\|$ = norma o magnitud del vector u .

Cabe recordar que en un espacio euclidiano, de entre todos los vectores en la dirección de u , hay uno que es el más próximo a v . A esto se le llama *mejor aproximación de v sobre u* , y es precisamente la proyección ortogonal $\text{Proy}_u(v)$. (Neila, 2014, p. 15.).



Cuadro 1

Varianza promedio explicada por I frente a Y_1

Dimensión p	$n = 100$			$n = 2456$			$n = 10000$		
	I	Y_1	Y_1/I	I	Y_1	Y_1/I	I	Y_1	Y_1/I
3	2.525	2.526	1.00038	2.484	2.485	1.00048	2.475	2.476	1.00045
4	3.277	3.278	1.00020	3.189	3.190	1.00026	3.181	3.182	1.00027
5	3.995	3.995	1.00012	3.869	3.869	1.00016	3.856	3.856	1.00015
6	4.739	4.740	1.00007	4.576	4.577	1.00008	4.551	4.551	1.00010
7	5.481	5.481	1.00004	5.224	5.224	1.00006	5.207	5.207	1.00007
8	6.190	6.190	1.00003	5.891	5.891	1.00004	5.858	5.859	1.00005

Fuente: cálculos propios; para la comparación de las varianzas explicadas, los vectores de ponderaciones deben ser previamente normalizados, es decir, tener longitud unitaria.

5. Resultados de la simulación

Se simuló un experimento en el que se generaron variables correlacionadas distribuidas de forma aleatoria, para diferentes tamaños de población. El experimento se repitió 500 veces mediante la elaboración de una rutina en lenguaje de programación *R* (*R Core Team*, 2013).

En el cuadro 1 se observa que la varianza explicada por el nuevo índice (I) es casi la misma que la correspondiente a la primera componente prin-

cipal, Y_1 . De hecho, el cociente de la información retenida por esta componente respecto al nuevo índice es muy cercano a la unidad.

6. Resultados en la práctica

Para el análisis de resultados prácticos, se utilizaron los nueve indicadores básicos del índice de marginación del Consejo Nacional de Población (CONAPO, 2013) aplicados a las 32 entidades federativas del país, para 1990, 1995, 2000, 2005 y 2010.

Cuadro 2

Continúa

Estadísticas descriptivas, 1990-2010

Estadística	Año	I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8	I_9
Promedio	1990	12.54	38.88	22.37	13.12	20.36	57.87	21.07	37.42	62.72
	1995	10.79	24.25	14.35	7.40	14.84	64.91	17.13	34.65	64.26
	2000	9.47	29.48	10.55	5.03	10.49	46.31	14.47	33.55	51.65
	2005	8.36	23.78	5.88	2.72	9.59	41.17	11.11	31.43	46.08
	2010	6.83	20.20	3.95	1.98	8.05	36.80	6.32	30.61	39.50
Desviación estándar	1990	6.93	10.14	13.68	7.91	12.61	7.55	12.51	17.62	9.31
	1995	6.04	5.61	12.19	5.27	10.47	7.50	10.58	17.09	11.30
	2000	5.43	8.83	7.70	3.50	8.68	7.92	10.68	16.88	13.29
	2005	5.00	7.80	5.50	1.86	8.12	7.60	9.06	16.57	15.12
	2010	4.24	6.82	3.88	1.35	7.33	6.81	4.76	16.47	11.87
Coefficiente de variación	1990	55.26	26.09	61.17	60.29	61.91	13.04	59.40	47.10	14.85
	1995	55.97	23.12	84.90	71.11	70.54	11.56	61.76	49.32	17.58
	2000	57.35	29.94	72.94	69.57	82.70	17.11	73.81	50.33	25.73
	2005	59.80	32.81	93.47	68.44	84.69	18.46	81.57	52.71	32.82
	2010	62.07	33.76	98.34	67.87	91.00	18.50	75.27	53.81	30.06

Estadísticas descriptivas, 1990-2010

Estadística	Año	I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8	I_9
Correlación media ^a	1990	0.818	0.814	0.769	0.720	0.751	0.751	0.777	0.785	0.651
	1995	0.797	0.693	0.649	0.677	0.736	0.733	0.764	0.747	0.756
	2000	0.837	0.817	0.694	0.745	0.767	0.747	0.790	0.773	0.762
	2005	0.817	0.792	0.561	0.685	0.733	0.698	0.768	0.730	0.750
	2010	0.805	0.782	0.529	0.591	0.740	0.682	0.743	0.710	0.729

^a Entre cada variable y el resto (se utilizó la media cuadrática).

I_1 = % de población analfabeta de 15 años o más de edad.

I_2 = % de población de 15 años o más de edad sin primaria completa.

I_3 = % de ocupantes en viviendas sin drenaje ni escusado.

I_4 = % de ocupantes en viviendas sin energía eléctrica.

I_5 = % de ocupantes en viviendas sin agua entubada.

I_6 = % de viviendas con hacinamiento.

I_7 = % de ocupantes en viviendas con piso de tierra.

I_8 = % de población que vive en localidades menores a 5 mil habitantes.

I_9 = % de población económicamente activa que percibe hasta 2 salarios mínimos.

Fuente: cálculos propios con base en los nueve indicadores de marginación del CONAPO.

En el cuadro 2 se muestran algunas estadísticas descriptivas para los indicadores considerados. Ahí se observa que, en promedio, los indicadores de rezago social tienden gradualmente a la baja al transcurrir el tiempo, es decir, experimentan una mejora, pero según el coeficiente de variación, el beneficio del desarrollo entre las entidades no es equitativo; por lo tanto, el nuevo índice sintético tendrá que reflejar esta situación en términos numéricos.

Por su parte, las mayores dispersiones las presenta el indicador correspondiente a la población radicada en localidades rurales; en contraste, las menores se registran donde no disponen de energía eléctrica; por ejemplo, para el 2010, la varianza del primero es casi 150 veces más grande en relación con el segundo, lo cual significa que las varianzas de los indicadores son heterogéneas, a pesar de estar medidos en la misma escala (en porcen-

Índices y niveles de rezago social por entidad federativa, 1990-2010.
Primera componente principal (Y_1) frente al nuevo índice (I) proyectado a 1990

Primera componente principal (Y_1)					Ent.	Nuevo índice (I) proyectado a 1990					Niveles de rezago				
1990	1995	2000	2005	2010		1990 ^a	1995	2000	2005	2010	1990	1995	2000	2005	2010
2.36	2.33	2.25	2.32	2.32	07 Chis.	52.4	42.3	32.3	21.5	16.9	6	5	4	3	2
2.06	1.82	2.08	2.13	2.15	20 Oax.	49.7	38.7	31.2	20.7	16.3	5	4	4	3	2

Índices y niveles de rezago social por entidad federativa, 1990-2010.
Primera componente principal (Y_1) frente al nuevo índice (Z) proyectado a 1990

Primera componente principal (Y_1)					Ent.	Nuevo índice (Z) proyectado a 1990					Niveles de rezago				
1990	1995	2000	2005	2010		1990 ^a	1995	2000	2005	2010	1990	1995	2000	2005	2010
1.75	1.88	2.12	2.41	2.53	12 Gro.	47.0	39.1	31.4	21.9	17.6	5	4	4	3	2
1.17	0.99	0.88	0.75	0.66	13 Hgo.	41.8	32.8	23.4	14.6	11.2	5	4	3	2	2
1.13	1.12	1.28	1.08	1.08	30 Ver.	41.5	33.7	26.0	16.1	12.7	5	4	3	2	2
0.83	0.79	0.72	0.64	0.71	21 Pue.	38.8	31.4	22.4	14.1	11.4	4	4	3	2	2
0.75	0.75	0.72	0.66	0.56	24 SLP	38.1	31.1	22.4	14.2	10.9	4	4	3	2	2
0.57	0.59	0.30	0.16	0.10	32 Zac.	36.5	29.9	19.7	12.0	9.3	4	3	2	2	1
0.52	0.66	0.66	0.46	0.47	27 Tab.	36.0	30.5	22.0	13.3	10.6	4	4	3	2	2
0.48	0.77	0.70	0.57	0.43	04 Camp.	35.6	31.2	22.3	13.8	10.4	4	4	3	2	2
0.40	0.79	0.38	0.43	0.42	31 Yuc.	35.0	31.4	20.2	13.2	10.4	4	4	3	2	2
0.36	0.39	0.45	0.46	0.53	16 Mich.	34.7	28.5	20.7	13.3	10.8	4	3	3	2	2
0.21	0.13	0.08	0.10	0.06	11 Gto.	33.3	26.6	18.3	11.7	9.2	4	3	2	2	1
0.16	-0.19	-0.11	-0.14	-0.26	22 Qro.	32.8	24.4	17.1	10.6	8.0	4	3	2	2	1
0.01	0.00	-0.11	-0.02	0.05	10 Dgo.	31.5	25.7	17.0	11.2	9.1	4	3	2	2	1
-0.04	-0.23	-0.19	-0.14	-0.15	29 Tlax.	31.1	24.1	16.6	10.7	8.4	4	3	2	2	1
-0.13	0.05	0.06	0.19	0.12	18 Nay.	30.2	26.1	18.1	12.1	9.4	4	3	2	2	1
-0.14	-0.21	-0.10	-0.15	-0.26	25 Sin.	30.1	24.2	17.1	10.6	8.1	4	3	2	2	1
-0.19	-0.22	-0.36	-0.33	-0.42	23 Q. Roo	29.7	24.2	15.4	9.8	7.5	3	3	2	1	1
-0.46	-0.54	-0.36	-0.44	-0.27	17 Mor.	27.3	21.9	15.5	9.4	8.0	3	3	2	1	1
-0.60	-0.73	-0.60	-0.62	-0.55	15 Méx.	26.0	20.5	13.9	8.5	7.1	3	3	2	1	1
-0.61	-0.57	-0.69	-0.69	-0.72	28 Tamps.	26.0	21.7	13.3	8.3	6.5	3	3	2	1	1

Índices y niveles de rezago social por entidad federativa, 1990-2010.
Primera componente principal (Y_1) frente al nuevo índice (I) proyectado a 1990

Primera componente principal (Y_1)					Ent.	Nuevo índice (I) proyectado a 1990					Niveles de rezago				
1990	1995	2000	2005	2010		1990 ^a	1995	2000	2005	2010	1990	1995	2000	2005	2010
-0.76	-0.70	-0.69	-0.73	-0.78	06 Col.	24.7	20.7	13.4	8.1	6.3	3	3	2	1	1
-0.77	-0.59	-0.76	-0.77	-0.82	14 Jal.	24.6	21.5	12.9	7.9	6.1	3	3	2	1	1
-0.86	-0.84	-0.76	-0.75	-0.70	26 Son.	23.7	19.8	12.9	8.0	6.6	3	2	2	1	1
-0.87	-0.76	-0.78	-0.68	-0.52	08 Chih.	23.7	20.3	12.7	8.3	7.2	3	3	2	1	1
-0.89	-1.04	-0.97	-0.96	-0.91	01 Ags.	23.5	18.3	11.5	7.1	5.8	3	2	2	1	1
-0.97	-0.82	-0.80	-0.71	-0.68	03 BCS	22.8	19.9	12.6	8.2	6.6	3	2	2	1	1
-1.05	-1.16	-1.20	-1.14	-1.14	05 Coah.	22.0	17.5	10.0	6.3	5.1	3	2	2	1	1
-1.35	-1.25	-1.27	-1.25	-1.14	02 BC	19.4	16.8	9.6	5.8	5.1	2	2	1	1	1
-1.38	-1.47	-1.39	-1.33	-1.38	19 NL	19.1	15.2	8.8	5.5	4.2	2	2	1	1	1
-1.69	-1.71	-1.53	-1.50	-1.48	09 DF	16.3	13.5	7.9	4.7	3.9	2	2	1	1	1

^a Las observaciones están ordenadas con base en esta columna, en forma descendente.

Niveles de rezago: <10 = 1; [10,20) = 2; [20,30) = 3; [30,40) = 4; [40,50) = 5; >50 = 6.

Fuente: cálculos propios con base en los nueve indicadores de marginación del CONAPO.

taje), pero los recorridos son diferentes de forma notoria. Esto, junto con el hecho de que conceptualmente se establece la misma importancia para cada variable, conlleva a que el índice compuesto se genere a partir de las variables estandarizadas respecto a su desviación estándar, es decir, que se utilice la matriz de correlaciones en lugar de la de covarianzas.

Al final de este mismo cuadro se pueden ver las correlaciones medias entre cada indicador y el resto, donde se observa que, en general, son *altas* para los cinco años, por lo que, dado el traslape de información, es posible construir índices sintéticos que den cuenta de manera resumida de la situación multivariada prevaleciente en los conjuntos de datos originales.

En el cuadro 3 se presentan los resultados obtenidos tanto por la primera componente principal (Y_1) como por el nuevo índice (I), así como los niveles de rezago social a partir de este último; las observaciones están ordenadas de manera descendente conforme al nuevo índice para 1990.

Los límites de los intervalos de valores que corresponden a los niveles de rezago fueron establecidos de manera arbitraria, pero razonada, al observar la situación que prevalece en el año inicial (1990), esto es, no se utilizó ningún procedimiento de análisis de conglomerados; se mantuvieron fijos a lo largo del tiempo con el fin de observar los cambios de nivel entre las unidades de observación al pasar de un año a otro.

El nuevo índice aparece transformado al considerar la proyección ortogonal al año base 1990, una vez que es calculado para cada año; por ejemplo, los vectores de ponderaciones unitarios para 1990 y 1995, respectivamente, son:

$$u = (.499, .339, .238, .385, .252, .421, .262, .188, .295)'$$

$$v = (.472, .442, .191, .460, .252, .350, .258, .156, .239)'$$

Para encontrar la magnitud de la proyección del índice de Aguascalientes, de 1995 a 1990, el procedimiento es como sigue:

Primero se determina el vector que corresponde a la magnitud del índice en la dirección de v :

$$w = (18.6) * (.472, .442, .191, .460, .252, .350, .258, .156, .239)'$$

$$= (8.8, 8.2, 3.5, 8.6, 4.7, 6.5, 4.8, 2.9, 4.4)'$$

Así, conforme a lo expresado en el apartado 4, la magnitud del índice en la dirección de u para 1990 es:

$$w \cdot u = 18.3$$

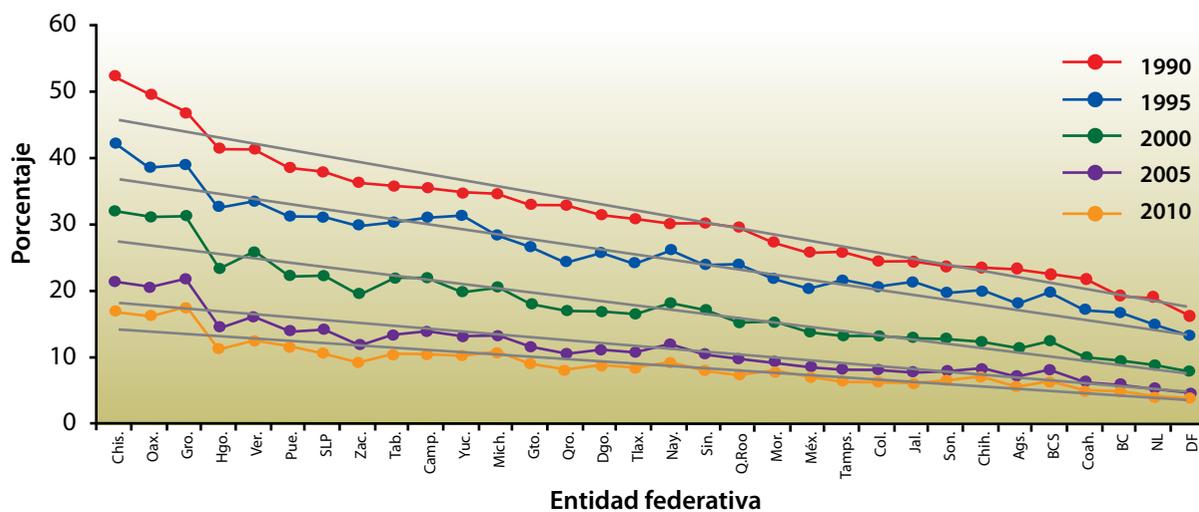
En este mismo cuadro destaca que el ordenamiento coincide con el de Y_1 (de hecho, las varianzas explicadas por ambos procedimientos también coinciden hasta del orden de centésimas y se ubican cerca de 80%); sin embargo, obsérvese como sus valores no son fáciles de interpretar: algunos son positivos y otros, negativos; en el tiempo suben y bajan, además de que no están en la escala de las variables originales ni poseen un cero absoluto. Esto es consecuencia principal de la forma en que se estandarizaron los indicadores básicos, ya que fueron expresados como desviaciones respecto a sus propias medias divididas entre sus correspondientes desviaciones estándar (Cortés y Vargas, 2011, p. 362).

En contraste, con el nuevo índice, estas limitaciones son esclarecidas. En particular, se observa que el rezago social disminuye a lo largo del tiempo para todas las entidades federativas, reflejando así el comportamiento de las variables originales, en una escala también de cero a cien en términos porcentuales.

Con base en los resultados del nuevo índice, en la gráfica 1 se ilustra la disminución gradual del rezago social en el tiempo; se observa que, en

Gráfica 1

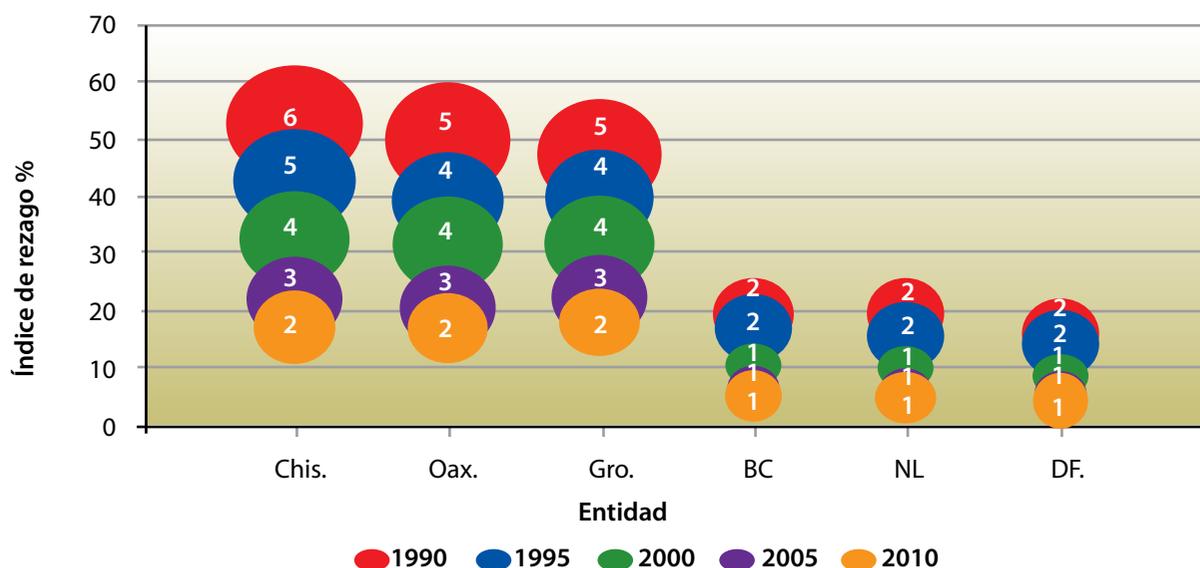
Índices de rezago social por entidad federativa, 1990-2010 (ordenadas conforme a I para 1990)



Gráfica 2

Índices y niveles de rezago social para entidades extremas, 1990-2010

(ordenadas conforme a *I* para 1990)



general, la brecha entre las entidades federativas es cada vez más cerrada.

Mientras tanto, en la gráfica 2 se destaca que las tres entidades con mayor rezago social son Chiapas, Oaxaca y Guerrero; en el otro extremo se ubican el Distrito Federal, Nuevo León y Baja California.

7. Conclusiones

El nuevo índice compuesto refleja la situación que prevalece en las unidades de observación, según las variables originales consideradas, esto es, permite un ordenamiento de las unidades de observación y mide los cambios aproximados en el tiempo.

Dicho índice está dado en la misma escala de medición que las variables de insumo (en particular, si éstas expresan un porcentaje); entonces, el índice tendrá una variación entre cero y 100.

El nuevo índice es fácil de calcular, de interpretar y, por lo tanto, también sin problemas para comunicarlo o explicárselo al usuario no especializado.

En contraste con la primera componente principal, el nuevo índice tiene una interpretación intuitiva, ya que al considerar el promedio de la correlación entre cada variable y el resto, de manera inmediata se asocia el mayor valor con la importancia de la variable correspondiente, y es ahí donde debe ponerse especial interés, de manera tal que invertir en ella para abatir su rezago, si fuera el caso, indirectamente se estaría mejorando la situación de las otras variables y, por lo tanto, beneficiando a la población objeto de análisis, por ser el fin último de las políticas sociales.

8. Consideraciones finales

La metodología para generar el índice tiene sentido cuando las variables son numéricas y están correlacionadas; de preferencia, medidas en la misma dirección, es decir, todas deberán ser de bienestar, o bien, de rezago. Por lo tanto, el índice es de uso generalizado a cualquier fenómeno con estas características de medición.

El nuevo índice sintético es con fines exploratorios y descriptivos; se aplica a datos poblacionales.

Su aplicación con datos muestrales y la consideración de dependencia temporal para realizar inferencia estadística quedan fuera del alcance de este trabajo, aunque pueden ser motivo de discusiones académicas y de estudios posteriores con mayor grado de especificidad.

Cuando se utiliza la matriz de varianzas y covarianzas los coeficientes que deben normalizarse son: c_i/S , donde c_i denota la covarianza y se define de manera similar como en el caso de r_i , pero ahora el promedio se obtiene sobre p ; S es la suma de los c_i . De esta forma,

$$I = \frac{c_1x_1 + c_2x_2 + \dots + c_px_p}{c_1 + c_2 + \dots + c_p}$$

donde,

$$c_i = \sqrt{\frac{c_{i1}^2 + c_{i2}^2 + \dots + c_{ip}^2}{p}}, \text{ para } i = 1, 2, \dots, p$$

Por último, con el propósito de facilitar el cálculo de indicadores compuestos y enriquecer el análisis, ya se cuenta con una aplicación informática en lenguaje de programación R, donde también es posible observar la relación entre variables provenientes de diversas fuentes, formar los factores inmersos y seleccionar aquél relevante para la construcción del índice sintético de interés; esto mediante la presentación y vinculación de gráficas multivariadas, como es el caso de las Biplot y Trellis, así como de las gráficas dinámicas, pero también con la aplicación de técnicas estadísticas recientes. En la actualidad, se está analizando la forma de incorporar elementos geográficos, como: ríos, carreteras, distancias y ubicación de los centros de salud. El sistema se encuentra en su versión beta y se espera que, en breve, el usuario pueda generar su propio índice en línea; para conocer más acerca de esto, el lector puede comunicarse con el autor del presente artículo.

Fuentes

- Campos, Nila. *Álgebra Lineal. Espacio Euclídeo*. España, Universidad de Cantabria. Consultado en: http://personales.unican.es/camposn/espacio_euclideo.pdf el 15 de mayo de 2013.
- Consejo Nacional de Población. Índices de marginación 1990, 1995, 2000, 2005 y 2010. México, consultados en: www.conapo.gob.mx/es/CONAPO/Indices_de_Marginacion el 11 de abril de 2013.
- Cortés, F. y D. Vargas. "Marginación en México a través del tiempo: a propósito del índice de CONAPO", en: *Estudios Sociológicos*. Vol. XXIX, núm. 86, mayo-agosto, 2011, pp. 361-387. Distrito Federal, México, El Colegio de México.
- Flury, B. N. "Common principal components in k groups", en: *Journal of the American Statistical Association*. 79, 1984, pp. 892-898.
- _____. "Two generalizations of the common principal component model", en: *Biometrika*. 74, 1987, pp. 59-69.
- Forni, Mario, Marc Hallin, Marco Lippi y Lucrezia Reichlin. "The generalized dynamic factor model: identification and estimation", en: *The Review of Economics and Statistics*. Noviembre del 2000.
- Geyne, F. y R. Hernández. "Análisis longitudinal de la marginación en México (1990-2005)", en: *XXII Foro Nacional de Estadística*, AME. 2007.
- INEGI. Censos y conteos de población y vivienda, 1990-2010. México, consultados en: www.inegi.org.mx/est/contenidos/proyectos/ccpv/default.aspx el 26 de noviembre de 2012.
- Jackson, J. Edward. *A User's Guide To Principal Components*. New York, John Wiley & Sons, Inc., 1991.
- Johnson, Richard A. y Dean Wichern. *Applied Multivariate Statistical Analysis*. New York, Prentice-Hall, Inc., 1982.
- Lavit, Ch. *Analyse Conjointe de Tableaux Quantitatifs*. París, Masson, 1988.
- Lavit Ch. y C. Roux. *Manual del método Statis*. Ponencia en ISUP, París, 1982.
- OCDE. *Handbook on constructing composite indicators: methodology and user guide*. París, Francia, 2008. Consultado en: <http://www.oecd.org/std/42495745.pdf> el 19 de marzo de 2013.
- Peña, Daniel. *Análisis de datos multivariantes*. McGraw-Hill Interamericana de España, 2003.
- R Core Team. *R: A language and environment for statistical*. 2013. Consultado en: www.R-project.org/
- Vences, José. *Estadística multivariada. Análisis de factores*. México, Instituto de Educación de Aguascalientes, 1999.
- Watson, Mark W. y Robert F. Engle. "Alternative algorithms for the estimation of dynamic factors, MIMIC, and varying coefficient regression models", en: *Journal of Econometrics*. 23, 1983, pp. 385-400.