ESTADÍSTICA ESPAÑOLA
núm. 108, 1985, págs. 57 a 64

Un análisis del indicador usado en el "Fondo de Compensación Interterritorial"

por JESUS BASULTO SANTOS
Facultad de Ciencias Económicas
y Empresariales
Universidad de Sevilla

RESUMEN

En este trabajo se define una distancia que permite medir la aproximación entre dos Comunidades y la posición de cada Comunidad respecto de una Comunidad imaginaria. A partir de la distancia se analiza la representatividad del indicador usado en el Fondo de Compensación Interterritorial, proponiendo un indicador alternativo de asignación.

Palabras clave: Comunidad Autónoma; Distancia; Representatividad.

1. INTRODUCCION

El Fondo de Compensación Interterritorial consiste en repartir una cierta cantidad, C, entre las 19 Comunidades Autónomas, asignando a la Comunidad i–ésima (i=1, 2,. . ., 19) la cantidad C. Z (i), siendo Z (i) el valor del indicador Z sobre dicha Comunidad.

El indicador Z se define como,

$$Z = I_1.w_1 + I_2.w_2 + I_3.w_3 + I_4.w_4 \qquad (1)$$

donde I_1, I_2, I_3 e I_4 son indicadores que enseguida definiremos y $w_1 = 0.7$, $w_2 = 0.2$, $w_3 = 0.05$ y $w = 0.05$ son pesos que miden, respectivamente, la importancia de cada uno de los indicadores I_j, j=1, 2, 3, 4.

El indicador I_1 se define como,

$$I_1 = Z_1 / T_1$$

siendo $Z_1 = (10^6 / R) . (R_m / R) . P$

y donde P es la población, R la renta per cápita, R_m la renta mínima entre todas las Comunidades y T_1 es el total de Z_1 para las 19 Comunidades. Notemos que si Z_1 se define por,

$$Z_1 = P / R^2$$

entonces no se modifica el indicador I_1; así R_m no tiene ninguna influencia en el indicador I_1.

El indicador I_2 se define como,

$$I_2 = Z_2 / T_2$$

siendo $Z_2 = Q.S$

y donde S es el saldo migratorio, Q es una variable que toma el valor -1 si S es menor que cero y el valor 0 si S es mayor o igual que cero, y T_2 es el total de Z_2 para las 19 Comunidades.

El indicador I_3 se define como,

$$I_3 = Z_3 / T_3$$

siendo $Z_3 = J. (TP - TMP)$

y donde TP es la tasa de paro de cada Comunidad, TMP es la tasa de paro en España, J es una variable que toma el valor 1 si TP supera a TMP y el valor 0 en otro caso, y T_3 es el total de Z_3 para las 19 Comunidades.

Por último, el indicador I_4 es la proporción de superficie de cada Comunidad.

Para más detalle sobre las últimas definiciones, puede consultarse la publicación: Indicadores Estadísticos Regionales (I.N.E., 1984). También, en nuestro trabajo no incluímos el indicador que incorpora, en la asignación, el alejamiento entre las islas y la península.

Notemos que por (1), el 70% de C se reparte proporcionalmente al indicador Z_1, el 20% al Z_2, el 5% al Z_3 y el 5% restante al indicador Z_4.

Si consideramos la matriz X con 19 filas (las Comunidades) y 4 columnas (los indicadores), donde el elemento de la fila i–ésima y de las columnas j–ésima, x_{ij}, es igual a I_j (i); siendo I_j (i) el valor del indicador I_j sobre la Comunidad i–ésima, entonces podemos valorar la aproximación entre las Comunidades i e i' por la distancia,

$$d_{w'}(i, i') = \sqrt{\sum_{j=1}^{4} w_j \cdot (x_{ij} - x_{i'j})^2} \tag{2}$$

donde, otra vez, los pesos w_j (j=1, 2, 3, 4) permiten dar mayor importancia a unas columnas, de la matriz X, más que a otras en el cálculo de dicha distancia.

Ahora, si consideramos que el indicador Z representa la fila i–ésima de la matriz X por el valor Z (i), podemos pensar que dos Comunidades i e i' con distancia, $d_{w'}(i, i')$, pequeña (grande) deben tener asignaciones parecidas (muy diferentes); es decir $|Z(i) - Z(i')|$ debe ser pequeño (grande). Así, un primer objetivo de nuestro trabajo es analizar hasta qué punto $d_w(i, i')$ está bien descrita por $|Z(i) - Z(i')|$, para todo i e i'.

También, si consideramos una Comunidad imaginaria, k, tal que $I_j(k) = 0$ (j=1, 2, 3, 4), entonces podemos asignar la cantidad C, a la Comunidad i–ésima, proporcionalmente a

$$d_w(i, k) = \sqrt{\sum_{j=1}^{4} w_j \quad x_{ij}^2} \tag{3}$$

y así, un segundo y último objetivo de nuestro trabajo será relacionar éste criterio con el basado en el indicador Z.

2. EVALUACIÓN DE LA REPRESENTATIVIDAD DEL INDICADOR Z.

Si cada fila $x_i = (x_{i1}, x_{i2}, x_{i3}, x_{i4})$, i=1, 2,. . ., 19, de la matriz X se considera como un punto o vector del espacio vectorial R^4 (donde R es el conjunto de los números reales), cuyas componentes son sus coordenadas en la base canónica (e_1=(1, 0, 0, 0), e_2=(0, 1, 0, 0), e_3=(0, 0, 1, 0) y e_4=(0, 0, 0, 1)), y si en R^4 consideramos la distancia definida en (2), entonces podemos planteamos el hallar un vector m_1, de norma unidad según (2), y tal que minimice la expresión,

$$\sum_{i=1}^{19} d_w^2(x_i, x_i^*) \tag{4}$$

donde x_i^* es la proyección ortogonal, según la norma definida por (2), sobre el vector m_1.

Es fácil ver que (4) es igual a,

$$\| X - X^* \|_{I,w}^2 = \text{traza } (I.(X-X^*).W.(X-X^*)') \tag{5}$$

donde la fila *i*–ésima de la matriz X^* es x_i^* y la matriz W es una matriz diagonal definida por W=diagonal (w_1, w_2, w_3, w_4), es decir en su diagonal principal tiene los pesos w_j, j=1, 2, 3, 4.

Ahora, el problema recogido en (4) equivale a: Hallar una matriz X^*, de rango unidad, que minimice (5) entre todas las matrices de rango unidad.

La solución de este último problema es (C. Eckart and G. Young, 1936; A.S. Householder and Gale Young, 1938) la matriz $X^* = d_1.n_1.m_1'$; donde a partir de la descomposición singular generalizada de la matriz X en,

$$X = N.D.M' \quad ; \quad N'.N = I \quad y \quad M'.W.M = I$$

siendo D = diagonal (d_1, d_2,. . ., d_r), r es el rango de X, resulta que n_1 y m_1 son las primeras columnas de las matrices N y M, respectivamente, y d_1 es el mayor valor singular de la matriz X.

Ahora, la fila *i*–ésima, x_i' de la matriz X puede representarse por la cantidad $x_i'.W.m_1$, que es la proyección ortogonal de x_i sobre m_1. Así, podemos definir el indicador B mediante,

$$B = I_1.w_1.m_1(1) + I_2.w_2.m_2(2) + I_3.w_3.m_3(3) + I_4.w_4.m_4(4)$$

donde $m_1(j)$ es la *j*–ésima componente del vector m_1.

Para la matriz X del año 1984, a partir de aplicar un algorítmo de potencias, hemos obtenido para d_1, n_1 y m_1 los siguientes valores:

d_1 = 0,3611

n_1' = (0'7881, 0'0649, 0'0598, 0'01903, 0'09091, 0'02408, 0'265, 0'2172, 0'234, 0'139, 0'2048, 0'3073, 0'1416, 0'0546, 0'0175, 0'09981, 0'00872, 0'007067, 0'00579)

$$m_1 = (0'94508,\ 1'1790,\ 1{,}13519,\ 0'804147)$$

y así el indicador B se escribe como,

$$B = I_1.w_1.(0'94508) + I_2.w_2.(1'1790) + I_3.w_3.(1'13519) + I_4.w_4.(0{,}804147)$$

En la siguiente tabla comparamos los valores que toman, sobre cada Comunidad, los indicadores Z y B.

	Z	B		Z	B
1	0'2818	0'2846	10	0'0535	0'0503
2	0'0251	0'0234	11	0'0729	0'0739
3	0'0219	0'0216	12	0'1078	0'1106
4	0'0073	0'0068	13	0'0542	0'0511
5	0'0334	0'0328	14	0'0221	0'0207
6	0'0089	0'0086	15	0'0068	0'0063
7	0'0948	0'0958	16	0'0360	0'0360
8	0'0772	0'0784	17	0'0034	0'0031
9	0'0874	0'0848	18	0'0026	0'0025
			19	0'0021	0'0021

donde no se observan diferencias relativas apreciables entre ambos indicadores.

La bondad de ajustar X por X* se define por,

$$\left(1 - \frac{\| X - X^* \|^2_{I \cdot w}}{\| X \|^2_{I \cdot w}}\right) . 100$$

que para nuestros datos resulta un valor de 89'3%.

También, al ser X'.W.X = (N.D) . (N.D)', resulta que la distancia (2) entre las Comunidades i e i' es igual a la distancia euclídea entre las filas i e i' de la matriz N.D. Si ahora aproximamos la matriz X por la matriz $X^* = d_1.n_1.m'_1$, resulta que X'.W.X puede ser aproximada por $(d_1.n_1)$. $(d_1.n_1)'$, es decir por la matriz $(X.W.m_1)$. $(X.W.m_1)'$; ya que $X.W.m_1 = d_1.n_1$. Entonces, podemos aproximar la distancia d_w (i, i') por $|B(i) - B(i')|$ o por $|Z(i) - Z(i')|$.

En la tabla siguiente se compara la distancia $d_w(i,i')$ con la distancia $| Z(i) - Z(i') |$ (que aparece entre paréntesis) para algunos pares i e i' de Comunidades.

	1	3	6	9	15	18
2	0'2641 (0'2567)	0'0176 (0'0032)	0'0232 (0'0162)	0'0881 (0'0623)	0'0228 (0'0183)	0'0287 (0'0225)
5	0'2528 (0'2484)	0'0343 (0'0115)	0'0381 (0'0245)	0'0627 (0'0540)	0'0392 (0'0266)	0'0416 (0'0308)
10	0'2392 (0'2283)	0'0426 (0'0316)	0'0528 (0'0446)	0'0653 (0'0339)	0'0543 (0'0467)	0'0586 (0'0509)
13	0'2385 (0'2276)	0'0451 (0'0323)	0'0552 (0'0453)	0'0662 (0'0332)	0'0569 (0'0474)	0'0608 (0'0516)
17	0'2825 (0'2784)	0'0191 (0'0185)	0'0059 (0'0055)	0'1025 (0'0840)	0'0042 (0'0034)	0'0022 (0'0008)
19	0'2834 (0'2797)	0'0283 (0'0198)	0'0070 (0'0068)	0'1035 (0'0853)	0'0063 (0'0047)	0'0006 (0'0005)

En la tabla se pueden observar grandes diferencias relativas entre dichas distancias; sobre todo cuando las distancias entre las Comunidades son pequeñas. Esto último se debe a que el indicador Z supone que el vector óptimo m_l es h'_l=(1, 1, 1, 1), resultando la matriz $(X.W.h).h'$ como la "mejor" aproximación a la matriz X, y siendo la bondad de ajuste menor o igual que 89'3%. Así, el resto de 11'7%, al menos, de no explicación del valor $\| X \|^2_{I,W}$ explica las diferencias relativas observadas en la última tabla.

Si la matriz X = N.D.M' se escribe como,

$$X = (N.D).M'$$

resulta que el valor x_{ij} (fila i–ésima y columna j–ésima) es el producto escalar, usual, de la fila i–ésima de la matriz N.D por la fila j–ésima de la matriz M. Ahora, con la aproximación de X por $d_l.n_l.m'_l$ resulta que la columna j–ésima de la matriz X puede representarse por $m_l(j)$, componente j–ésima de m_l. Así, en la aproximación de la matriz X por la matriz $(X.W.h_l).h'_l$, al ser h'_l=(1, 1, 1, 1) concluímos que las cuatro columnas de la matriz X deben ser iguales. Esto último no está de acuerdo con los datos recogidos en la matriz X.

Por último, como la distancia $d_w(i, k)$ puede aproximarse por B(i), se recogen en la siguiente tabla las comparaciones entre $d_w(i, k)$ y Z(i), i=1, 2,. . ., 19, así como las asignaciones que se obtienen cuando C es igual a 100.000 unidades monetarias.

	$d_{w'}$ (i,k)	ASIGNACIONES	Z (i)	ASIGNACIONES
1 Andalucía	0'2854	25611'03	0'2818	28180
2 Aragón	0'0309	2774'85	0'0251	2510
3 Asturias	0'0225	2023'18	0'0219	2190
4 Baleares	0'0084	761'32	0'0073	730
5 Canarias	0'0438	3933'44	0'0334	3340
6 Cantabria	0'0094	845'16	0'0089	890
7 Castilla-León	0'1050	9423'70	0'0948	9480
8 Castilla-La Mancha	0'0873	7839'94	0'0772	7720
9 Cataluña	0'1055	9471'16	0'0874	8740
10 C. Valenciana	0'0617	5544'31	0'0535	5350
11 Extremadura	0'0780	7005'99	0'0729	7290
12 Galicia	0'1181	10602'99	0'1078	10780
13 Madrid	0'0639	5740'43	0'0542	5420
14 Murcia	0'0253	2275'89	0'0221	2210
15 Navarra	0'0083	748'36	0'0068	680
16 País Vasco	0'0501	4501'19	0'0360	3600
17 Rioja (La)	0'0041	370'61	0'0034	340
18 Ceuta	0'0032	289'06	0'0026	260
19 Melilla	0'0026	237'31	0'0021	210

donde al ser la suma de la columna correspondiente a $d_{w'}$ (i,k) igual a 1,08372, resulta que las asignaciones son iguales a C.$d_{w'}$ (i,k)/1'08372, i=1, 2, ..., 19.

Como se puede observar en la última tabla, los criterios de asignación basados en $d_{w'}$ (i,k) y Z (i) presentan diferencias relativas significativas (superiores al 5%).

3. DISCUSION Y CONCLUSIONES

En el apartado anterior vimos que el indicador Z deja de explicar al menos el 11'7% de $\| X \|^2_{I,W'}$, y así no podemos igualar la distancia $d_{w'}$ (i,i') a la distancia $|Z(i) - Z(i')|$ para las Comunidades i e i'. También, se vió que el indicador Z está suponiendo que los cuatro indicadores (I_j : j=1, 2, 3, 4) toman el mismo valor sobre cada Comunidad y, por último, vimos que existen diferencias significativas entre las asignaciones basadas en los criterios $d_{w'}$ (i,k) y Z (i).

También, si se quiere tener un 100% de explicación para Z, se debería definir los indicadores I_j, j=1, 2, 3, 4, de tal forma que el rango de la matriz X fuese igual a la unidad. Ahora bien, al ser la suma de cada columna de la matriz X igual a la unidad, resultaría que sus columnas serían iguales en cada una de las filas, y así la introducción de los pesos no tendría ningún efecto sobre los resultados.

Concluímos, pues, en utilizar el criterio basado en la distancia d_{κ} (i,k), evitando criterios que se fundamenten en una *reducción de la información*. Así, nuestro trabajo, al igual que otros (J. Hernández, T. Mancha y J. E. Villena, 1981), pretende que, aceptando el concepto de distancia como criterio de asignación, se centren las críticas en la elección de los indicadóres y los pesos.

REFERENCIAS

ECKART, C. and GALE YOUNG (1936): "The approximation of one matrix by another of lower rank". *Psychometrika*, *1*, 211–218.

HERNÁNDEZ, J., T. MANCHA y J. E. VILLENA (1981): "Un fondo de compensación desvirtuado". *Estudios Regionales*, *8*, 17–35.

HOUSEHOLDER, A. S. and GALE YOUNG (1938): "Matrix approximation and latent roots". *Am Math. Monthly*, *45*, 165–171.

I. N. E. (1984): *Indicadores Estadísticos Regionales*.

SUMMARY

ANALYSIS OF THE INDICATOR USED IN THE "FUND FOR INTERTERRITORIAL COMPENSATION".

In this paper we are defined a distance to evaluate the approximation between one Community and another and to measure the position the each Community with respect to an imaginary Community. This distance is used to analyze the representativity of indicator utilized in the Fund for Interterritorial Compensation. As a consecuence, this article propose an alternative indicator for the F. I. C.

Key words: Autonomic Community; distance; representativity.

AMS 1970. Subject Classification: Primary 62P20,
Secondary 62H25.