

UNHA METODOLOXÍA PARA MEDI-LAS DISPARIDADES ECONÓMICO-SOCIAIS DAS COMUNIDADES AUTÓNOMAS E AS SÚAS NECESIDADES

AGUSTÍN RAMOS CALVO / JOSÉ CARLOS DE MIGUEL DOMÍNGUEZ
Departamento de Métodos Cuantitativos para a Economía e a Empresa
Facultade de Ciencias Económicas e Empresariais
Universidade de Santiago de Compostela

Recibido: 3 novembro 1997

Aceptado: 13 xaneiro 1998

Resumo: Proponse unha metodoloxía para cuantifica-lo reparto de fondos entre as distintas autonomías. Utilizando variables socioeconómicas que reflecten as disparidades existentes entre as distintas rexións españolas, é posible cuantifica-las necesidades de gasto de cada unha das provincias. Para isto, utilízase a análise de compoñentes principais (como método obxectivo) para delimita-las diferencias e, así, poder elaborar índices de necesidade de cada unha das autonomías usando a agregación dos resultados provinciais.

Palabras Clave: Solidariedade interrexional / Necesidades / Compoñentes principais / Índices.

A METHODOLOGY TO MEASURE SOCIO-ECONOMIC DISPARITIES OF THE AUTONOMOUS COMMUNITIES AND THEIR NEEDS

Summary: A methodology to value the revenue sharing among the different Autonomies is proposed. By means of socio-economic variables reflecting the existing disparities among the different Spanish regions, it is possible to assess the expenditure needs for each province. In this way, an analysis of the main components (as an objective method) is used to delimit differences and thus be able to work out the need indexes for each autonomy utilising the aggregation of the provincial results.

Keywords: Interregional solidarity / Needs / Principal components / Indexes.

INTRODUCCIÓN

No artigo 158 da Constitución, aparece a creación dun Fondo de Compensación Interterritorial (FCI) onde se recollan os gastos de investimento necesarios e tendentes a corrixir os desequilibrios económicos rexionais. En definitiva, trátase de facer efectivo o principio de *solidariedade interterritorial*.

O obxectivo básico deste estudo é proporcionar unha metodoloxía *distinta* á do custo efectivo para determina-los gastos públicos de cada comunidade. A razón de propoñer este método baséase en que o uso do custo efectivo conduciu —probablemente— a perpetua-las desigualdades existentes anteriores ó Estado das autonomías. Trátase, polo tanto, de analizar unhas regras de comportamento mediante as que cada goberno (e, en consecuencia, os cidadáns correspondentes) participe nos recursos comúns en función das diferencias entre as necesidades de gasto que

poden garanti-la igualdade dos servicios dos que se goza. Todo iso variará —como é obvio— en función da capacidade fiscal propia de cada rexión.

Anque se tratan en profundidade as cuestións relativas ás necesidades de cada autonomía e as barreiras de acceso ós servicios polos residentes de cada comunidade, non se tratará sobre o concepto e definición da *capacidade* de cada goberno autónomo para atende-las carencias citadas; en consecuencia, tampouco se intentará cuantificar esa capacidade nin as posibilidades dos gobernos citados para xerar ingresos propios.

Debe observarse que, no presente estudio, non se trata a cuestión referente ó importe total do Fondo de Compensación, de modo que sexa *suficiente* para os fins que se lle asignan. Non obstante, establécese información sobre as condicións necesarias para articula-los mecanismos de decisión que permitan conseguir unha asignación territorial do gasto público que sexa, á vez, *eficiente e solidaria*.

Antes de concreta-lo método, cabe citar que o FCI inclúe, ademais da solidariedade interrexional, o financiamento dos servicios transferidos. Esta cuestión (que en certo modo complementa á anterior) chega a suscitar problemas de incompatibilidade: non parece fácil conxuga-la necesidade de investimentos novos nos servicios transferidos do Estado coa solidariedade necesaria para que todos reciban os mesmos servicios independentemente do lugar de residencia. De feito, abonda con contempla-la compensación transitoria dos últimos anos para comprobar que a necesidade de garantir recursos adicionais —que faciliten a suficiencia financeira das comunidades— pode afondar no sistema de financiamento, pero non contribúe ó principio de solidariedade interrexional nunha dimensión adecuada. Insistindo niso: nos últimos anos, a redistribución cara ás rexións menos desenvolvidas acentúase, pero a custa doutras comunidades que gozan de escaso desenvolvemento. E todo porque as máis desenvolvidas teñen transferencias de servicios nos que se debe investir (Fernández Caínzos, J.; Álvarez Corbacho, X.; Fernández Leiceaga, X.; Caramés Viéitez, L. (1996), pp. 77-92).

SOBRE AS NECESIDADES DE GASTO

Enténdese que dúas comunidades teñen as mesmas necesidades de gasto nun determinado servicio cando as posibilidades de *acceso* dos seus habitantes a ese servicio é o mesmo. Trátase, polo tanto, de estudia-las posibilidades de que os cidadáns gocen dos mesmos servicios públicos, independentemente da súa situación económica. Ese principio de igualdade de acceso pode interpretarse en termos económicos de exceso de demanda, no sentido de que é igual a probabilidade de atende-la demanda dun individuo calquera. Unicamente cabe sinalar que:

- O concepto de necesidade que se está manexando aparece como comparación entre comunidades e é, por conseguinte, un concepto relativo.

- É fundamental estimar as necesidades correspondentes á igualdade de acceso ós servizos públicos *primarios* (por exemplo, a sanidade) que non admiten factores discriminadores entre as comunidades. No caso de servizos *secundarios* (como, por exemplo, o apoio á actividade económica) estimaranse as condicións de igualdade a partir da demanda actual, que non é a mesma en tódolos casos.

Á hora de medi-lo *exceso de demanda* como estimación da igualdade de acceso aparecen —entre outras— as seguintes dificultades:

- A idea de demanda de servizos leva implícito o suposto de que o individuo se comporta na súa elección dun modo "habitual". Claramente, non é a situación do caso que nos ocupa (a modo de exemplo, nos servizos sanitarios é o médico o que case sempre decide polo paciente).
- Os servizos primarios son, moi a miúdo, gratuítos. De aí que, anque se pode interpretar a demanda en función do que o consumidor está disposto a ofrecer en termos de tempo e enerxía (en lugar do prezo), xa non é fácil obter información sobre o oferente do servizo porque o que o consumidor proporciona xa non chega a aquel. Abondando sobre o tema, a demanda que resulta desta situación ten unha interpretación distinta dende o punto de vista do benestar.
- Finalmente, é preciso resaltar que a maior dificultade estriba na identificación das variables que han determina-la oferta e a demanda anteriores. Non adoitan existir variables que reflictan esa oferta independentemente da demanda e —incluso— en certos casos a demanda estará determinada pola oferta en modo perverso (así por exemplo, unha redución da oferta sanitaria pode crear a longo prazo un grande aumento da demanda polos efectos que a carencia poida ter).

Tampouco parece fácil nin factible formular unha estimación da demanda acudindo ós orzamentos dos gobernos locais e delimitando as variables que os determinan; a vida das comunidades é suficientemente curta como para que esa situación estea reflectida adecuadamente. Todo isto comporta acudir ó concepto de *necesidade* e (na bibliografía existente sobre o tema) sóese recorrer á intuición e postula-la existencia de indicadores que permitan medila. En todo caso, é importante precisa-lo concepto para saber exactamente qué supostos se están facendo e deriva-las propiedades do reparto dende o punto de vista do benestar.

Por todo iso, cabe insistir en que se usa un concepto *comparativo*, no sentido de estudia-la necesidade como algo que xorde ó comparar dúas comunidades con carácter análogo e observar por qué unha delas presta máis e (ou) mellores servizos cá outra. Ademais, como as comunidades teñen asignadas competencias en distintos sectores como a sanidade, as obras públicas, etc., falar de necesidade pode implicar un certo grao de *invasión desas competencias*. Como consecuencia, en lugar de definir unha política sobre eses servizos, o que se procura é tratar de

cuantifica-la necesidade buscando indicadores —o máis neutrais posible— que representen un denominador común a tódalas políticas razoables de servizos públicos.

Para rematar, é preciso non esquece-lo efecto que a iniciativa privada poida ter sobre a provisión deses servizos. Por exemplo, se nunha comunidade a iniciativa privada ofrece vivenda accesible ós cidadáns, estes adquiren a vivenda por esa vía e os indicadores que se estimen reflectirán unha necesidade baixa; en definitiva, a comunidade citada pode quedar discriminada por esa iniciativa. Ademais, será preciso ter en conta se os gobernos autonómicos financian esa iniciativa, porque nese caso segue existindo un índice de necesidade que supón no diminuí-los recursos recibidos dende fóra. Non é doado, pois, separar eses efectos da iniciativa privada sobre os indicadores estimados.

REPARTO EN FUNCIÓN DAS NECESIDADES

A necesidade relativa da poboación esixe unha distribución de recursos que pode ser xustificada do seguinte modo:

- Existen dous tipos de barreiras (o "prezo" ó que se aludía anteriormente) ás que o cidadán debe enfrontarse para recibi-los servizos. Unhas de *necesidade común*, dependentes do tamaño da comunidade na que se recibe o servizo, e outras de *necesidade adicional*, baseadas nas características socioeconómicas específicas de cada poboación. Claramente, un maior tamaño da poboación implica unha meirande necesidade de accesos, pero —ademais— ante a mesma masa poboacional, tamén aumenta a dificultade proporcionalmente a certas características singulares de cada sociedade (así por exemplo, un maior índice de anciandade supón un maior volume de recursos empregados para atender esa poboación e, de rebote, menos posibilidades para o resto).
- Pode parecer que simplifica-lo problema reducíndoo a estes dous tipos de necesidades supón actuar lonxe da realidade; pero en favor dese tipo de actuación cabe argüír que, anque a realidade non é así, ese efecto simplificador supón que outras circunstancias (aínda que existen) *non distorsionan* o comportamento entre comunidades. Poñamos un caso concreto —posiblemente o máis claro—: se a oferta de servizos é moi diferente, o reparto que resulta das consideracións anteriores non garantirá a igualdade dos accesos a eses servizos. Agora ben, se dúas comunidades coas mesmas necesidades (común e adicional) parten dun capital social diferente, a que dispón de menor capital ten ó seu favor que a súa necesidade de recursos para amortizalo pode ser tamén menor e, así, pode formular unha política máis agresiva cos recursos que lle quedan para tratar de alcanzar (co tempo) os servizos da outra. En definitiva, anque o anterior pode non ser bastante, supón satisfacer unha condición necesaria para conseguila igualdade.

- Por último, aínda que non se determine o que é necesidade común e específica, é fácil interpreta-las consecuencias redistributivas derivadas de valoracións distintas de cada tipo de necesidade: canta maior ponderación se lles conceda ás necesidades comúns, máis progresivo será o reparto porque as comunidades pobres recibirán máis. Pola contra, as características diferenciais supoñen desviar fondos cara ás comunidades máis ricas. No límite que supón diminuí-lo peso das necesidades comúns e aumenta-lo das específicas, chegaríase a considerar que todas elas teñen a mesma importancia e chegar así a un reparto que sería directamente proporcional á poboación.

Todo o anterior esixe insistir nas seguintes puntualizacións:

- 1) Xa que as diferencias de nivel obtidas para as distintas comunidades constitúen a base do cálculo de provisión adicional de recursos, postular de antemán unhas barreiras de acceso e a súa ponderación esixiría o uso dunha análise factorial da situación. Neste estudio utilízase a *análise das compoñentes principais*, sendo a diferenza fundamental co anterior que as ponderacións (e, en definitiva, a das barreiras) serán as que resulten dos indicadores neutrais seleccionados.
- 2) A análise é fundamentalmente operacional, o que significa que:
 - a) O tamaño das "barreiras de acceso" é o reflectido polos indicadores.
 - b) Ese tamaño non é único, senón un entre moitos.
 - c) Non se formula a existencia dun *modelo verdadeiro*, senón que o que se persegue é conseguir, con cada medición, un perfil do fenómeno que se estudia.
- 3) Como non se formula a existencia do modelo, non ten sentido buscar unha compoñente común ós indicadores que recolla as correlacións con aqueles e, en consecuencia, o peso que cada indicador poida ter na "necesidade relativa".
- 4) Por último, somos conscientes do rexeitamento que adoitan sufrir en certos círculos este tipo de análises, xa que as variables observadas son simples indicadores duns parámetros que se definen *a posteriori*. Ademais, non é fácil interpreta-los estimadores obtidos ou é preciso facelo cunha maior precaución ca nun modelo de ecuacións simultáneas. A favor desta formulación pódese razoar que, ademais de que formula-lo modelo de antemán supón invadir competencias das comunidades, ó noso xuízo presentaría grandes dificultades formular ese modelo de maximización do benestar unido a unhas restriccións iniciais sobre as compoñentes principais como un conxunto de ecuacións estruturais.

CONCEPTO DE COMPOÑENTES PRINCIPAIS (CP)

Suposta unha variable $x=(x_1, x_2, x_3, \dots, x_p)'$ con media 0 e Σ matriz de varianzas-covarianzas, con descomposición espectral $\Sigma \equiv \Gamma \Lambda \Gamma'$, sendo $\Gamma=(\gamma_1, \gamma_2, \gamma_3, \dots, \gamma_p)$ a

matriz de autovectores e $\Lambda=(\lambda_1, \lambda_2, \dots, \lambda_p)$ a de autovalores ordenados: $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$, denomínase transformación de compoñentes principais á obtida facendo $Y = \Gamma' x$. O vector $Y_j = \gamma'_j x$ chámase *j-ésima compoñente principal de x*.

Esas CP gozan de propiedades interesantes tanto alxebraica coma estatisticamente, que poden resumirse en :

- 1) $E(y_j) = 0$; $V(y_j) = \lambda_j$, $\forall j = 1, 2, \dots, p$.
- 2) $\text{Cov}(y_i, y_j) = 0$; $\forall i, j$ distintos.
- 3) Ningunha combinación lineal normalizada de x ten maior varianza que λ_1 , a varianza da primeira compoñente principal.

En definitiva, as CP do vector x son aquelas combinacións lineais dese vector que recollen a maior varianza. Son centradas, as súas varianzas son os autovalores correspondentes e, finalmente, non teñen correlación entre si. Dese modo, constitúen uns novos eixes de coordenadas que son ortogonais para explica-las variacións do vector x .

Unha vez introducidas as CP e as súas características fundamentais, é preciso sinala-la súa grande aplicación no estudio concreto de cuestións nas que se maneja unha gran cantidade de información; tanto polo número de variables utilizadas, como pola cantidade de observacións de cada unha desas variables. Concretamente, se se analizan p variables representativas do fenómeno estudado e N observacións de cada unha desas variables, é posible reduci-la dimensión do espacio \mathfrak{R}^N de observacións, como a do espacio \mathfrak{R}^p de variables en estudio. Para iso, abonda con prescindir das CP menos significativas, que son as correspondentes ós autovalores máis pequenos. Ademais, a análise de CP constitúe un método obxectivo de elección das variables máis significativas do fenómeno; esa obxectividade débese a que a elección das variables se determina unicamente en función da varianza (autovalor correspondente) explicada por cada unha delas. Por outro lado, na redución da dimensión polo uso de CP non desaparecen as variables iniciais, senón as combinacións lineais daquelas que menos varianza total explican. Dese modo, recóllense tódolos efectos que unha variable pode ter en cada observación particular e non se elimina a posibilidade de recoller ese efecto noutras circunstancias.

A grandes trazos, o proceso sinalado anteriormente supón estudar algunhas cuestións delicadas respecto da análise efectuada; a modo de exemplo cítanse algunhas circunstancias que determinan o estudio:

- 1) É posible efectuar tests de hipóteses para determinar de qué número de CP se pode prescindir tendo en conta a distribución da suma dos últimos autovalores. Para N suficientemente grande, os autovalores son variables normais e, por conseguinte, esa suma é unha χ^2 polo que é posible construír unha F de Snedecor se se considera a suma total de autovalores.

- 2) Para rexeitar algunha ou algunhas das compoñentes, tamén debe terse en conta a presenza de *esfericidade* nas CP. En caso de que isto ocorra, debe terse en conta o criterio de rexeitalas ou admitilas todas á vez pero, en ningún caso, prescindir só dalgunha delas.
- 3) Non parece tan delicada a situación que se pode crear cando algunha das variables iniciais é *redundante* nas compoñentes non rexeitadas. Aínda no caso de que o fose en todas elas, o seu efecto será casi nulo e, por conseguinte, a súa presenza non ten por qué restarlles robustez ós estimadores. Posiblemente, presentaríase esa circunstancia se se prescindise da variable.

SOBRE O TRATAMENTO DAS VARIABLES UTILIZADAS

Para conseguir-los obxectivos propostos, pódense considerar tódalas variables que se estimen representativas inicialmente e as súas correspondentes observacións en cada unha das provincias españolas. Ademais, os grupos que compoñen a mostra son as comunidades autónomas. Xa que o número de observacións relativas a cada unha destas é pequeno (e tendo en conta a distribución asintótica de autovalores e autovectores), optouse polo método de agregación dos resultados provinciais. Queda aberta a posibilidade de poder efectua-lo estudio utilizando CP para cada un dos grupos que constitúen as comunidades e analiza-la existencia de compoñentes comúns a tódolos grupos, que puideran recoller unha porcentaxe significativa da varianza explicada.

Entre os moitos problemas tanto operativos coma conceptuais que o estudio presentado carrega e que foi preciso resolver, cabe cita-los seguintes:

- a) Non se parte dunha definición apriorística das necesidades que se deben cubrir, xa que, aínda que existen deficiencias que cubrir universalmente aceptadas (como poden se-las relativas ás necesidades sanitarias), non é tan claro determinar se outras deben terse en conta ou non e o seu grao de importancia (como por exemplo, os investimentos necesarios para mellorar unha actividade económica determinada).

En lugar de fixar criterios que cuantifiquen as cuestións anteriores, optouse por buscar unha medida de necesidade recollendo a información de tódalas variables iniciais mediante as CP e ponderando o peso de cada variable mediante o coeficiente que se obtén da matriz estrutural (matriz de factores) das compoñentes.

- b) Aínda que non se usa a análise factorial en todo o exposto, é coñecido o feito de que as CP son un caso particular daquela (véxase, por exemplo, J.T. Scott). Prefírese utiliza-las CP porque a análise factorial esixe que as comunilidades estean fixadas de antemán; isto esixiría prefixa-las variables no estudio, contravindo a intención expresada no apartado anterior.

- c) É preciso face-lo cálculo das CP mediante o uso do paquete informático SPSS e así se formulan (ver Aparicio Pérez) dificultades que, en esencia, supoñen que o que se acha son combinacións lineais das variables iniciais nas que non entran tódalas variables, senón un subconxunto delas expresado mediante unha "comunalidade" que varía entre 0 e 1. Para evitar esta situación, é preciso introducirle restricións ó programa e conseguir que a comunalidade sexa 1 en tódolos casos. É o único modo que coñecemos para obter con seguridade as CP. Aínda que tamén existe a posibilidade de facer rotacións sobre os eixes principais obtidos, evitouse utilizalas porque iso desvirtúa as propiedades esenciais de non-correlación entre as CP.
- d) Tódolos resultados relativos á distribución dos estatísticos usados para os tests de hipóteses sobre as CP e as súas varianzas (λ_i) están baseados na hipótese de normalidade das variables iniciais. Xa que moitas das variables usadas non son nin sequera simétricas, débese conseguir "normalizalas" utilizando calquera das transformacións coñecidas (cuadrática, raíz cadrada, logarítmica). Na maioría dos casos, ese proceso conduce a resultados inaceptables. É aconsellable utilizar transformacións do tipo Box-Cox, co que se consegue un alto grao de normalidade tanto en cada unha das variables coma no seu conxunto.
- e) Unha vez que se atoparan as compoñentes e as puntuacións que nelas posúen os individuos (provincias), pódese optar por obter a información de cada un destes sumandos nas puntuacións correspondentes a aquelas CP que recollen unha determinada proporción significativa de varianza (técnica usada en investigacións de mercado para establece-los pesos relativos dos factores obtidos na análise de CP para os datos dunha enquisa), ou sumar esas puntuacións ponderadas pola varianza correspondente. Claramente, débese elixir a segunda posibilidade, xa que o método exposto separa as *combinacións lineais* segundo a súa habilidade para explica-la varianza común dos datos orixinais: sería un contrasentido utilizalo método e, a continuación, dálle-la mesma importancia a todas elas.
- f) Utilizar unha gran cantidade de variables no estudio pode parecer repetitivo; máis aínda se existe un alto grao de correlación entre elas. Se o estudio se fai de forma global, o anterior conduce a resultados inaceptables. O caso é distinto cando se distinguen diversas *categorías* de necesidade, como se expoñerá máis adiante.
- g) Finalmente, a elaboración de índices globais de necesidade precisa dunhas novas ponderacións para as distintas categorías (sanidade, vivenda, etc.) que, agora si, non son medibles obxectivamente. A importancia que a cada unha delas lles conceden as autonomías nos seus orzamentos é distinta; pero, en todo caso, é independente dos acordos ós que se chegou sobre financiamento coa administración central. Considérase oportuno ponderar coas medias correspondentes desas asignacións autonómicas.

VARIABLES INICIAIS

No noso estudio utilizáronse ata 38 variables iniciais para analizar conxuntamente as necesidades. Seguindo os criterios de Bosch e Escribano (1988) e outros de elaboración propia, esas variables iniciais agrúpanse por categorías seguindo os criterios seguintes:

◆ *Sanidade*

- 1) Indicadores de morte e enfermidades: mortalidade total e infantil, morbilidade.
- 2) Necesidade de atención médica: anciáns que viven sós, dependencia demográfica, poboación necesitada e natalidade.
- 3) Dificultades de asistencia: vivendas sen auga corrente, entidades por km², densidade de poboación, estadías e gastos hospitalarios.

◆ *Vivenda e urbanismo*

- 1) Esfuerzo de reconstrución e mellora: vivendas sen servicios hixiénicos, sen auga corrente ou sen rede de sumidoiros. Consumo de enerxía eléctrica e iluminación pública.
- 2) Necesidades de ordenación do territorio, do turismo e das persoas de paso: prazas hoteleiras, presión demográfica, variación da poboación e saldo migratorio interior.
- 3) Necesidades asociadas a certo grao de desenvolvemento: porcentaxe de ocupados na industria e na construción.

◆ *Transporte e comunicacións*

Ademais dun indicador da dotación telefónica, considéranse:

- 1) Necesidades de mantemento e estado da rede viaria: quilómetros de estrada por entidade de poboación, número de accidentes mortais, quilómetros de autoestradas e autovías, transporte marítimo e aéreo. Total de vehículos matriculados.

◆ *Servicios sociais*

Recóllense, en xeral, variables indicativas relativas ós grupos de poboación susceptibles de melloras en servicios elementais mediante axudas institucionais, como poden ser anciáns que viven sós, parados, persoas menores de 5 e maiores de 70 anos, saldo migratorio interior, iluminación pública, índice de xuventude, densidade de poboación.

◆ *Lecer e cultura*

Recóllense, en xeral, indicadores de necesidades de centros culturais e recreativos segundo distintas categorías: parados, bilingüismo, cambios poboacionais dos

últimos anos, densidade de poboación, porcentaxe de poboación en núcleos rurais, índice de xuventude.

◆ *Medio natural*

Entre a gama de variables socioeconómicas, seleccionáronse aquelas que recollen un alto grao de expansión demográfica, o que conduce a unha maior necesidade de rede de sumidoiros, depuración de augas, tratamento de residuos, etc.: ocupados na industria, enfermidades de declaración obrigatoria, vehículos vellos, vivendas sen servizos elementais, prazas hoteleiras e ocupación, variación da poboación.

◆ *Educación*

Os criterios utilizados neste apartado son moito máis variables ca nos casos anteriores. En xeral, téndese a substituí-la análise factorial por algún índice obxectivo que reflecta a necesidade de investimentos en ensino. O método que se seguiu é o mesmo, considerando variables como: índice de persoas que estudian respecto ás que —teoricamente— deberían estudar, número de entidades de poboación por superficie, bilingüismo e densidade de poboación.

Malia a gran cantidade de variables citadas, non se esquecen moitas outras que tamén son significativas. A modo de exemplo, pódense citar algúns casos clarificadores:

- Non se dispón de datos actualizados da porcentaxe de casas vellas.
- A porcentaxe de poboación que reside na capital de provincia (Zaragoza, por exemplo).
- Elaboración dun índice que recolla a influencia dunhas boas vías de comunicación para zonas densamente poboadas ou cunha industria preponderante.

METODOLOXÍA PARA OS INDICADORES INICIAIS

Para conseguir unha certa homoxeneidade entre indicadores que miden magnitudes tan dispares, procúrase presentalos en forma de índices. Nos que non é posible, considéranse índices adecuados que reflecten a tendencia que se desexa cuantificar. Non obstante, anque o anterior poida parecer suficiente, o método definitivo consiste en que tódalas variables estean normalizadas de modo que sexan os valores máis distantes da media os que determinen as CP definitivas.

Tendo en conta que os valores erráticos alteran fortemente a distribución das medias xerais e que as distribucións dos autovalores son asintóticas, considéranse os valores das variables en tódalas provincias, e non nas 17 autonomías xa que esta

última situación suporía unha perda significativa de graos de liberdade nas distribucións.

Á parte das transformacións usuais (xa citadas) para conseguirla normalidade das variables, usáronse transformacións Box-Cox do tipo:

$$x^{(\lambda)} = \begin{cases} ((x+m)^\lambda - 1) / \lambda & \text{se } \lambda \neq 0 \\ \lg(x+m) & \text{se } \lambda \equiv 0 \end{cases} \quad x+m \geq 0$$

que, claramente, engloban a tódalas anteriores. Con estas transformacións e supoñendo que as variables transformadas $x_i^{(\lambda)}$ están idénticamente distribuídas como $N(0, \sigma^2)$, a función de verosimilitude dos datos orixinais toma a forma:

$$f(x, \lambda) = (2\pi\sigma^2)^{-n/2} \exp(-\sum (x_i^{(\lambda)} - \mu)^2 / 2\sigma^2) (\prod x_i^{\lambda-1})$$

na que o xacobiano non depende de λ e μ e, por conseguinte, os estimadores max-verosímiles (para λ fixo) son:

$$X(\lambda) = \sum x_i^{(\lambda)} / n \quad \text{e} \quad \sigma^2 = \sum (x_i^{(\lambda)} - X)^2 / n$$

polo que, se é x^* a media xeométrica dos x_i , daquela (agás constantes) o valor máximo da función de verosimilitude é:

$$L_{\max}(\lambda) = -(n/2) \lg \sigma^2 + n \lg x^{*(\lambda-1)} = -(n/2) \lg \sigma_z^2; \text{ sendo } z_i^{(\lambda)} = x_i^{(\lambda)} / x^{*(\lambda-1)}$$

Seguindo os traballos de Box-Cox e P.J. Bickel, pódese elixir-lo valor de λ que maximiza esa función para transforma-los datos iniciais. Apelando á precaución necesaria para non modificar excesivamente a información contida nos datos iniciais, pódese achar un intervalo de confianza para λ utilizando que:

$$L_{\max}(\lambda^*) - L_{\max}(\lambda) \leq \chi^2_{1,\alpha}$$

na que se está usando, implicitamente, que baixo a hipótese nula, $\lambda = \lambda_0$ a expresión:

$$2((L_{\max}(\lambda) - L_{\max}(\lambda_0)))$$

se distribúe como unha χ^2 con 1 grao de liberdade.

Utilizando a debida cautela no uso desas transformacións e seguindo resultados obtidos en distintas simulacións (J.C. Estévez e outros) os criterios de tratamento das variables iniciais son os seguintes.

- Transfórmanse aquelas variables cun coeficiente de asimetría que supera a cota de 0,5. Abonda con observar exemplos de variables que non superan esa cota para comprobar que o método é adecuado.
- Unha vez obtido o estimador max-verosímil para λ e o intervalo de confianza correspondente (naquelas variables que superan a cota anterior), transfórmase a variable utilizando o valor do extremo dese intervalo que é máis próximo a 1 (polo que a transformación resulta se-la máis "suave" entre tódalas posibles).
- Se nunha variable que non cumpre o primeiro criterio, o intervalo de confianza para $\hat{\lambda}$ contén o valor 1, préfírese deixala no seu estado inicial.

A nosa experiencia (op. anterior citada) aconsella actuar deste modo, xa que cando as variables iniciais son "groseiramente" normais, a transformación Box-Cox non altera a estrutura fundamental da información, mentres que esa alteración se agrava canto máis próxima á normalidade está a variable.

Finalmente, é necesario sinalar que, como é ben coñecido, a normalidade marginal das variables estudadas non garante a normalidade multivariable correspondente. Coas transformacións anteriores conséguense, cando menos, cumprir unha condición necesaria para que os estimadores obtidos sexan significativos. Dese modo, sempre estará xustificada utiliza-los tests necesarios de isotropía, redundancia, etc. que garanten a obxectividade da información obtida nos estimadores.

ELABORACIÓN DUN ÍNDICE DE NECESIDADE

A matriz R de correlacións entre as variables iniciais correspondentes ás distintas categorías consideradas é simétrica e definida positiva (con probabilidade 1). Dese modo, a súa diagonalización espectral pode ser escrita na forma $R=C'DC$, sendo C a matriz ortogonal de autovectores e $D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_s)$ a dos autovalores correspondentes; s indica o número de CP usadas logo de reduci-la dimensión.

Como no programa SPSS de análise factorial a descomposición anterior se fai na forma $R = A A'$, resulta $A = C D^{1/2}$, sendo $D = \text{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots, \sqrt{\lambda_s})$, polo que é A ($s \times s$) a matriz que define as ponderacións (é dicir, as correlacións) de cada variable dentro de cada factor. Lémbrese que a matriz A posúe as seguintes propiedades:

- 1) A suma de cadrados dos elementos de cada fila é o autovalor λ_i .
- 2) A suma de cadrados dos elementos de cada columna é 1, por ser C ortogonal.

3) O produto interior entre dúas filas (respectivamente columnas) é o coeficiente de correlación das variables relativas a esas filas (respectivamente é 1, por seren correlación os factores).

Por outro lado, é preciso ter en conta que os factores F (da análise factorial) se definen na forma:

$$X = A F$$

ou ben

$$F = (A'A)^{-1} A'X$$

e como é D a matriz diagonal de autovalores, a igualdade conduce á expresión seguinte:

$$F_i = (f_{i1}, f_{i2}, f_{i3}, \dots, f_{iN}) = \lambda_i^{-1} (\sum a_{ki} x_{k1}, \sum a_{ki} x_{k2}, \dots, \sum a_{ki} x_{kN}).$$

Unha vez obtidas as CP (que en definitiva son os eixes canónicos do hiper-elipsoide de datos) obtéñense as *puntuacións dos individuos* neses eixes; é dicir, as coordenadas das variables iniciais neses eixes ortogonais. Para iso, abonda efectua-lo produto

$$(C_{ik}) = (L_{ii}) \times (X_{ik}),$$

sendo C a matriz (s, N) de correlacións citadas, L a matriz desas compoñentes e X a matriz de datos orixinais. Obviamente, s é o número de compoñentes e N o de provincias consideradas.

Xa se citou o feito de que en numerosos estudos de mercado son estas as puntuacións usadas para obte-las conclusións. Nós consideramos máis adecuado utiliza-las *puntuacións ponderadas* polos autovalores correspondentes e iso supón efectua-lo cálculo

$$(D_{ik}) = \text{Diag} (\lambda_1, \lambda_2, \dots, \lambda_k) (C_{ik})$$

e esa expresión en termos da matriz A toma a forma

$$D = \text{Diag} (\lambda_1^{3/2}, \lambda_2^{3/2}, \dots, \lambda_k^{3/2}) (A'A)^{-1} A'X.$$

¹ Nótese que se $R=AA'$, daquela a estrutura do espectro de R mantense en calquera rotación determinada por unha matriz ortogonal T , xa que $(TA)'(TA) = A'T'TA = D$.

Dependendo de que se usen puntuacións C ou D , os índices correspondentes varían de forma significativa; en todo caso, o seu cálculo efectúase calculando a media aritmética simple das puntuacións. Cabe sinalar, ademais, que cando se opta por prescindir dalgunhas das CP para quedarse coas que recollen unha porcentaxe acumulada (por exemplo, de 90%) de varianza, esa media calcúlase tendo en conta a suma dos autovalores correspondentes ás compoñentes seleccionadas.

TRANSFORMACIÓN FINAL DOS ÍNDICES OBTIDOS

Xa que o cálculo das compoñentes esixe que os valores das variables estean tipificados, os índices obtidos toman valores positivos e negativos segundo que sexan maiores ou menores cá media. Obviamente, para poder interpretalos como índices de necesidade relativa, é preciso que todos eles sexan positivos. A necesidade de transformalos formula unha elección non só non é trivial, senón que afecta de modo substancial á sensibilidade dos parámetros distributivos que deben reflecti-las diferencias detectadas por provincias ou comunidades.

Chegados a este punto e antes de buscar unha transformación concreta, consideramos interesante sinalar que se entre dúas provincias non houbera maior diferenza cá masa poboacional, todo o estudio estaría de sobra. Os índices, precisamente, recollen as diferencias socioeconómicas entre elas. Así, é o momento de decidir "politicamente" o peso que se debe dar a esas diferencias, para que o reparto que dimana dos índices poida variar dunha forma ou doutra a cantidade asignada per cápita.

De entre outras moitas posibilidades e tendo en conta a súa versatilidade, óptase por unha transformación do tipo seguinte: Si é k un parámetro (que denominamos "*progresividade*"), a función de transformación utilizada é

$$g(\varepsilon) = \begin{cases} 1 + (k - 1)e^{\varepsilon} / 2 & \text{se } \varepsilon < 0 \\ 1 - (k - 1)e^{-\varepsilon} / 2 & \text{se } \varepsilon \geq 0 \end{cases}$$

Observarase que os valores da función g están comprendidos entre 1 e k e que a transformación permite que o índice correspondente a unha determinada provincia sexa k veces maior có doutra. O reparto será, así, máis ou menos progresivo segundo sexa maior ou menor o valor dese parámetro. Pode observarse que a función $g(\varepsilon)$ aumenta as diferencias dos valores próximos a 0, mentres que para valores moi baixos ou moi altos de ε o que consegue é aproximalos a 1 e k , respectivamente.

Finalmente, unha vez obtidos os índices provinciais relativos, obtéñense os índices para as comunidades mediante a suma ponderada daqueles, podendo utilizar

como ponderacións as poboacións relativas daquelas ou outro tipo de pesos que se considere máis conveniente.

CONSIDERACIÓNS FINAIS

- 1) Aínda que é desexable unha fórmula do reparto que sexa o máis simple posible, non é aconsellable (de cara a esa simplicidade) sacrificar variables cunha influencia nos indicadores que pode ser pequena, pero que a súa importancia nalgúns casos particulares resulta ser definitiva. Co método presentado, esa situación queda excluída.
- 2) Os índices obtidos son unha mostra de que os indicadores de necesidade non son únicos. Os seus valores dependen das variables iniciais elixidas, das compoñentes eliminadas e das ponderacións asignadas a elas; tamén a transformación final dos índices permite varia-la súa "progresividade". Nese sentido, pódese considerar subxectiva a elaboración dos índices a pesar da súa robustez. Non obstante, consideramos de utilidade a análise presentada por canto é unha medida directa das necesidades e, por conseguinte, serve como guía útil para chegar a un acordo razoable da distribución dos recursos entre as comunidades.
- 3) A pesar de todo o anterior, o método é *neutral* no sentido de que, fixadas as variables que dan lugar á elaboración do índice, a súa obtención está baseada en criterios de máxima varianza explicada para as CP e no da progresividade da transformación final.
- 4) Logo de comproba-los efectos que as variables teñen sobre as CP e, en definitiva, sobre os índices calculados, non é conveniente pensar nunha asignación de recursos a longo prazo. Esa asignación debe ser revisada continuamente tendo en conta novas variables e novas ponderacións que reflectan a importancia de cada unha das categorías de servicios.
- 5) Da análise de resultados obtidos, deducimos que —ante unha *necesidade común* a varias comunidades— non é adecuado un reparto proporcional á poboación desas comunidades, senón que deben se-las características específicas de cada unha das poboacións as que poñan de manifesto as necesidades adicionais. O comentario é especialmente válido no caso da renda: aínda que a renda é un bo indicador do grao de participación nos recursos que se reciben (de feito, é o que determina as comunidades beneficiarias do Fondo de Compensación), convértese nun indicador moi rudimentario se se usa para determina-las necesidades que intentamos cuantificar.
- 6) Aínda que nos índices elaborados coa metodoloxía exposta se recolle a influencia de demasiadas variables (circunstancia que pode desdebuxa-lo coñecemento estricto das necesidades) e esa tendencia pode aparecer repetida, non nos parece fácil nin aconsellable eliminar demasiadas variables no estudio.

BIBLIOGRAFÍA

- ANDERSON T.W. (1963): "Asimptotic Theory for Principal Components Analysis", *Annals of Mathematical Statistics*, vol 34, pp. 122-148.
- APARICIO PÉREZ (1988): "La difícil realización del Análisis de Componentes Principales mediante los programas estadísticos más difundidos en el mercado", *Estadística Española*, vol. 10, núm. 117.
- BICKEL, P.J.; DOKSUM, K.A. (1981): "An Analysis of Transformations Revisited", *JASA*, vol. 76.
- BOSCH-ESCRIBANO (1989): *Cinco estudios sobre financiación autonómica*. Madrid: Instituto de Estudios Fiscales.
- BOX, E.P.; COX, D.R. (1985): "An Analysis of Transformations", *Journal of the Royal Statistical Society*, vol. 26.
- BOX, E.P.; COX, D.R. (1982): "An Analysis of Transformations Revisited", *JASA*, vol. 77.
- ESTÉVEZ, J.C.; MIGUEL, J.C. DE; RAMOS, A. (1992): "Simulaciones en las transformaciones a la normalidad", *Congreso de ASEPELT*. Granada.
- FERNÁNDEZ CAÍNZOS, J.; ÁLVAREZ CORBACHO, X.; FERNÁNDEZ LEICEAGA, X; CARAMÉS VIÉITEZ, L. (1996): *Galicia en clave económica*, pp. 77-92.
- HILL, R.C.; FOMBY (1977): "Component Selection Norms for Principal Components Regression", *Biometrics*.
- MOLDES, E.; PUY, P. (1996): *La financiación de las Comunidades Autónomas*. Minerva.
- SCOTT, J.T. (1996): "Factor Analysis and Regression", *Econometrica*.
- TACKEUCHI, K. E OUTROS (1982): *The Foundations of Multivariate Analysis*. Wiley.
- WILKINSON, J.H. (1995): *The Algebraic Eigenvalue Problem*. Oxford: Clarendon Press.