

JOSE I. GURRIA GASCON

(Univ. Extremadura)

Cuando desbordados por una inabarcable matriz de datos y un tanto deslumbrados en principio por sus resultados, nos planteamos la aplicación de un análisis factorial en el marco espacial extremeño, no éramos conscientes de sus inconvenientes y limitaciones, a veces graves.

Sus resultados, aunque aceptables, no fueron los esperados debido a problemas de aplicación.

Ahora, la comunicación que presentamos es el resultado de largas meditaciones sobre esta serie de problemas de aplicación que limitan sus posibilidades. Esperamos con ello poder contribuir en alguna medida a un más correcto manejo y conocimiento de la mencionada técnica, en aras de alcanzar mejores resultados.

El objeto, pues, de la presente comunicación es el poner de relieve los problemas de aplicación e interpretación fundamentales que a nosotros concretamente nos han surgido. Pero no por ello y en absoluto pretendemos minimizar sus grandes posibilidades y ventajas, que por otra parte ya han quedado probadas en múltiples escritos y demostradas con sus resultados.

1.- CARACTERISTICAS, AMBITO DE APLICACION Y OBJETIVOS DEL ANALISIS FACTORIAL.

Conscientes de la complejidad de las interrelaciones que de

finen todo espacio geográfico, se confeccionó una matriz de datos con 111 municipios de la provincia de Cáceres y 37 variables, todas las que fue posible recopilar y cuantificar. Estas variables se pueden agrupar en tres conjuntos: variables físi- cas, variables demográficas y variables económicas, con especial incidencia en el sector agrario.

A esta matriz de datos se le aplicó un Análisis Factorial en Componentes Principales. El cálculo de los mismos se realizó por "Rutinas de Análisis de Varianza Multivariante". Y la tipi- ficación de los distintos municipios y comarcas por el cálculo de las Distancias de Mahalanobis.

Nuestro objetivo fundamental era definir y tipificar la mon- taña en sí y los municipios con economía de montaña, para su posterior estudio.

Para ello, incluimos en la matriz de datos todos aquellos municipios con algún porcentaje de su superficie por encima de los 500 m. (cota a partir de la cual los relieves ya destacan sensiblemente sobre su entorno, la penillanura). Y además, otro conjunto de municipios típicamente de penillanura, con el obje- to de definir la montaña con relación a su entorno y con el ob- jeto de que sirvieran como grupo "test" al análisis factorial, es decir, para comprobar si el análisis factorial los definía como tal. Junto a ellos y con la misma finalidad, los principa- les núcleos urbanos de la provincia.

Era evidente que cada uno de estos tres subsistemas habría de definirse por sus propias peculiaridades y por eso se codi- ficaron las variables típicas de montaña y también las más re- presentativas de la penillanura y de los núcleos urbanos.

Con ello, y aunque nuestro objetivo fundamental era la montaña, se podían definir y tipificar los distintos subsistemas regionales.

2.- LOS RESULTADOS.

La matriz de funciones discriminantes para las variables ya reducidas, aparecían constituyendo cuatro estructuras fundamentales: las dos primeras contenían nueve variables interrelacionadas entre sí cada una de ellas, y las dos segundas seis cada una. Todas estas funciones estaban jerárquicamente ordenadas, según la información que cada una aportaba a la definición del sistema, y distribuidas según una ley de composición interna. De acuerdo con esta ley, cada estructura se compone de una serie de funciones positivas, propias de las características de la penillanura, y otra serie de funciones negativas, propias de las características de la montaña. Es decir que, de acuerdo con ello, cada subsistema se definía por sus propias características, y por contraposición, por las propias de los otros, por las que no lo caracterizaban. Como indica FERNANDEZ GUTIERREZ, F (1978, pág. 93) "... cada factor puede tener dos polos que son dialécticamente opuestos... pero ambos polos integran y definen al factor... no son independientes... solos no son nada pero unidos, aunque opuestos, hacen un todo!"

En síntesis, los cuatro primeros factores, que venían definidos esencial y respectivamente por cada una de las cuatro estructuras mencionadas anteriormente, explicaban el 46'5% de la varianza.

Aparecían perfectamente individualizados los tres subsistemas, a pesar de explicar un porcentaje bastante bajo de la va-

rianza: la montaña, la penillanura y los núcleos urbanos. En la penillanura: los piedemontes, el área de riveros (de fuertes pendientes por el encajamiento del Tajo y sus afluentes), el regadío y la penillanura propiamente dicha. Y los núcleos urbanos, muy separados de los grupos anteriores, constituyendo un conjunto muy homogéneo.

Esta clasificación respondía perfectamente y con claridad a los esquemas tradicionales y a lo que era evidente.

El estudio posterior y detallado de estos conjuntos, tanto en las variables incluidas en la matriz de datos como en otras que no se pudieron cuantificar, demostraba la validez de estos resultados.

Sin embargo, un porcentaje tan bajo de la varianza invitaba y obligaba a la reflexión.

Para un mayor detalle de los planteamientos y resultados del análisis factorial que comentamos, remitimos a nuestra publicación "El paisaje de montaña en Extremadura (delimitación, economía y población) (1985).

3.- LOS PROBLEMAS DE APLICACION.

Los problemas de aplicación comienzan con la matriz de datos inicial, aunque las dificultades más importantes se deriven de la matriz de correlaciones.

Tendríamos que hablar en primer lugar de las deficiencias de las fuentes estadísticas, bien por defectos en su elaboración, bien por ausencia total o parcial, pero como son cuestiones de sobra conocidas y padecidas por todos nosotros, no haremos mayor hincapié. En nuestro caso, para la definición de la

montaña, se han echado en falta las variables climáticas y biogeográficas en general, por la ausencia de una auténtica red de estaciones meteorológicas, muy especialmente en altitud, y por la falta y dificultades de codificación de los suelos.

Hay que mencionar igualmente los siempre delicados trabajos de codificación de algunas variables de carácter cualitativo u ordinal. Aunque se pueden introducir matrices de ausencia-presencia o matrices de correlación de rangos (BROCARD, M PUMAIN, D et REY, V, 1977), siempre es comprometido y arriesgado, por cuanto que la correlación y el análisis factorial por lo tanto, son muy sensibles a los valores extremos. En este sentido, pues, hay que operar con todas las precauciones para no incurrir en graves anomalías.

En nuestro caso, existe una redundancia en la matriz de datos, en las cuatro variables de población activa, por cuanto que lógicamente son complementarias y han de estar por ello entre sí fuertemente correlacionadas. En este sentido, como indica el GROUPE DUPONT (1975, pág. 229), "chaque variable doit être passée travers trois cribes succesifs: a.- une variable doit mesurer un seul caractère, sans ambiguïté...; b.- si deux variables ont une signification trop voisine, elles peuvent par que fortement corrélées, déterminer à elles seules un axe secondaire, lorsqu'elles sont peu prises en compte par le premier facteur; c.- lorsque de fortes corrélations entraînent des groupes de variables à l'intérieur de la matrice, ceux-ci déterminent les axes".

Y esto es lo que ocurre con la citada redundancia, que viene a definir casi con exclusividad al segundo factor.

Uno de los problemas fundamentales es el del número total

de variables que constituyen la matriz de datos. Todas ellas, pocas o muchas, constituyen en cualquier caso el 100% de la varianza que tratan de explicar los distintos factores. Así pues, el análisis factorial considerado trata de explicar el 100% de uno o varios aspectos de una realidad mucho más compleja; el 100% de aquellos aspectos, y sólo de aquellos, que se hayan incluido en la matriz de datos.

El hecho de decir que explica mayor o menor porcentaje de la varianza, en principio no implica gran cosa. Es más, se puede manipular -y de hecho se hace de forma reconocida- para que explique un mayor porcentaje de la varianza.

Pasemos a la práctica del análisis factorial que nos ocupa. La matriz de datos, como se vió, consta de 37 variables. Las 37, que evidentemente no pueden ser todas las que configuran el 100% de ese complejo espacio montañoso, son sin embargo el 100% de "una realidad total" para el análisis factorial.

Y resulta que existen tantos factores como variables y todos explican algún porcentaje de esa varianza hasta el F_{35} . Pero desde el F_5 , ninguno aporta más del 2 ó 3 por ciento de la varianza, en su mayoría menos del 2%. Tenemos, por lo tanto, treinta factores que en su conjunto aportan casi el 50% de la varianza, pero individualmente no tienen ninguna importancia.

Si se eliminan, como de hecho se hace, las variables que que están menos correlacionadas y se dejan sólo las que se encuentran más correlacionadas, aumentará el porcentaje de la varianza explicado por los primeros factores. Es decir, se acumula en ellos lo explicado por los factores eliminados, aunque no sea la expresión correcta. Pero de esta manera, tendríamos una realidad más parcial, al menos distinta, con menos varia-

bles y manipulada: ¿Qué importaría entonces decir que en vez del 46'5%, los cuatro primeros factores explican casi el 80%?. Lo que se ha hecho no es sino reducir ese espacio a las estructuras resultantes más significativas.

Y no siempre conviene olvidar todas aquellas variables poco correlacionadas, pues muchas veces pueden aportar tanta información como las de mayores índices de correlación. Pero el análisis factorial consideraz casi exclusivamente a aquellas variables -como es lógico- que aportan mayor información a la definición de cada factor, las más interrelacionadas, y teniendo poco en cuenta a aquellas que presentan unos índices bajos.

Para una mejor explicación de este aspecto -uno de los problemas más serios del análisis factorial- es necesario recurrir a la correlación y a sus problemas de aplicación, ya que el análisis factorial se basa en primera instancia en la matriz de correlaciones.

Una variable que presente valores muy similares a lo largo de toda la muestra de municipios, no podrá estar interrelacionada con ninguna otra, pues sus valores ni aumentan ni disminuyen a medida que pueda hacerlo cualquier otra variable. Veamos un ejemplo:

Municipios	Variable "X"
1	10%
2	10
3	10
4	10
5	10
6	10

La variable "X", cualquiera que sea, nunca podrá estar

correlacionada con ninguna otra. Apenas será tenida en cuenta para la tipificación de los distintos municipios, así como tampoco en la definición previa del sistema. Sin embargo, nos está expresando una uniformidad, una homogeneidad, de esa variable a lo largo de toda la muestra de municipios y, en consecuencia, debiera de ser tenida en cuenta como una característica propia de todo el conjunto.

En este otro ejemplo, real (GURRIA GASCON, J.L., 1985₀), se puede constatar que las variables altitud y pendiente presentan un índice muy bajo (0'118), por cuanto que se trata de los municipios más montañosos. Es un conjunto, por lo tanto, compacto, muy homogéneo, similar en estas dos variables, y no se rían tenidas en cuenta apenas:

Municipios	Altitud	Pendiente	
1	98%	90%	
2	99	98	
3	98	95	
4	97	99	R=0'118
5	96	94	
6	98	97	

7	3	5	R=0'996

pero si incluimos un nuevo municipio con valores extremos, el índice aumenta hasta 0'996.

Hay que precisar de todo esto varios aspectos que consideramos de la máxima importancia:

- 1.- La correlación y el análisis factorial en consecuencia son muy sensibles a los valores extremos, que pueden introducir graves anomalías. Hay que operar por ello con sumo cuidado.

2.- Si estamos de acuerdo con BUCKLEY, W (1973, pág. 16), en que "la meta principal del movimiento de investigación de Sistemas Generales es, pues, perfilar esas semejanzas estructurales, y al mismo tiempo, distinguir las diferencias estructurales entre sistemas de tipos sustancialmente distintos", hemos de concluir que el análisis factorial no sirve para definir y caracterizar en sí mismos espacios y estructuras homogéneas; las variables que pudieran indicarlo apenas son tenidas en cuenta por su homogeneidad a lo largo de toda la muestra. El análisis factorial sólo diferencia espacios geográficos, no por las propias características y estructuras de cada uno de ellos, sino por contraposición a otros espacios distintos.

Si sólo se introduce un subsistema muy homogéneo, el porcentaje de la varianza explicado será muy bajo y se deberá a pequeñas peculiaridades; y, por otra parte, no aparecerá como un espacio homogéneo probablemente, si existen características muy contrastadas aunque no constituyan precisamente lo más representativo del subsistema o sistema en cuestión. Por ello, será necesario introducir en la matriz de datos subsistemas sustancialmente distintos y contrastados; en nuestro caso: la montaña, la penillanura y los núcleos urbanos. Y, aún en este caso, algunas variables (todas aquellas relacionadas con la vocación ganadera y forestal de la montaña) pueden aparecer como distintivas y diferenciadoras, como de hecho sucede, ya que la penillanura también tiene esta misma vocación aunque se deba a otras razones. En este ejemplo real, el sector forestal y la ganadería no están correlacionadas (presentan índices in-

feriores a $\pm 0'2000$ con todas las demás variables) con ninguna otra, pues sus valores son muy homogéneos a lo largo de toda la muestra de municipios. Si no están correlacionadas, no aportarán información y no tendrán peso en la definición y tipificación del sistema y de los distintos municipios. Apenas serán tenidas en cuenta, de donde se deriva que, a pesar de que la penillanura y la montaña tienen unos aprovechamientos fundamentalmente ganaderos y forestales, estos espacios no se cacterizarán por ello, sino por variables más específicas de cada uno de ellos. Esto mismo sucede con otras variables, como las demográficas, por insertarse ambos espacios en un contexto general de emigración y despoblación a nivel regional; la mecanización; las bajas rentas per cápita, etc.

En definitiva, las variables no distintivas casi no forman parte de las funciones discriminantes que definirán a los principales factores. Por el contrario, las más específicas de cada espacio serán las que aporten mayor información y tengan mayor peso: parcelas inferiores a una hectárea, altitud, pendiente, robledal, etc., en la montaña; encinar-alcornocal, tierras de labor, cultivos herbáceos, etc., en la penillanura.

A tenor de los resultados del análisis factorial, la montaña extremeña no encajaría en el contexto general de la montaña mediterránea en cuanto a sus aspectos y predominio de los aprovechamientos forestales y ganaderos. Sin embargo, un simple recorrido por el área de estudio demuestra todo lo contrario.

Todo ello implica que las distintas variables deberán matrizarse según su contexto más próximo y que el análisis factorial presente mejores resultados en espacios con característic

cas muy diferenciadas, teniendo en cuenta que variables que pueden ser distintas en un espacio más o menos reducido, pueden no serlo en otros espacios mayores o viceversa, con lo que las correlaciones se desfigurarían y con ellas también en el análisis factorial.

Las similitudes, por lo tanto, deberán resolverse a juicio y criterio del investigador, pues unas variables poco correlacionadas, no tenidas apenas en cuenta, pueden expresar similitudes y la caracterización fundamental de uno o varios subsistemas.

En síntesis, el análisis factorial define espacios geográficos de una forma parcial, sólo a través de sus estructuras distintivas, aunque no representen las características más definitivas de esos espacios. Más que definir espacios geográficos homogéneos, diferencia espacios geográficos distintos. Es, desde este punto de vista, una técnica "excepcionalista", ya que individualiza los distintos espacios por sus características, estructuras, excepcionales. No se podrá por ello teorizar ni buscar leyes generales a partir de sus resultados. Pero no es esta una afirmación que pretenda ser aseverativa, sino más bien un planteamiento a discutir.

En consecuencia, aunque el análisis factorial en componentes principales pueda ser una herramienta de gran utilidad, presenta problemas de aplicación e interpretación que pueden acarrear una desfiguración total o parcial de la realidad. Deberá prevalecer, por lo tanto y en todo momento, el criterio objetivo del investigador y el recurso a todo el material disponible y el estudio de campo.

BIBLIOGRAFIA

- BROCARD, M; PUMAIN, D et REY, V (1977): "L'analyse de données: traitements visuels et mathématiques" **L'Espace Géographique** VI, 4. Paris. Doin Ed.
- BUCKLEY, W (1973): **La Sociología y la Teoría Moderna de Sistemas**. Buenos Aires (2ª ed.). Amorrortu.
- ESTEBANEZ ALVAREZ, J. (1981): "Problemas de valoración e interpretación de los mapas mentales". **Anales de Geografía**, 1. Madrid. Universidad Complutense.
- FERNANDEZ GUTIERREZ, F. (1978): "Consideraciones metodológicas y experimentales del análisis factorial en Geografía". **Cuadernos Geográficos**, 8. Granada. Univ. de Granada.
- GRUPE DUPONT (1975): "La distance a la ville: essais d'analyses factorielles appliquées aux cas de Grenoble et de Montpellier". **L'Espace Géographique**, 4. Paris. Doin Editeurs.
- GURRIA GASCON, J.L. (1985): **El paisaje de montaña en Extremadura** (delimitación, economía y población). Cáceres. Servicio de Public. de la Univ. de Extremadura, Excmas. Diputaciones Provinciales y Junta de Extremadura.
- GURRIA GASCON, J.L. (1985): "La correlación lineal: precisiones prácticas y su funcionalidad en la determinación de las similitudes y diferencias de los espacios geográficos". **Norba**, V. Cáceres. Serv. Publicaciones de la Univ. de Extremadura.