

RASGOS BÁSICOS DEL SECTOR AGRARIO CASTELLANO-LEONÉS: UN NUEVO PLANTEAMIENTO ESPACIAL Y TEMPORAL. METODOLOGÍA STATIS

Ramón ALVAREZ ESTEBAN, María GÓMEZ RIOCEREZO,
Carmen HUERGA CASTRO, Eva VALLEJO PASCUAL
Universidad de León

1. INTRODUCCIÓN

El objetivo que pretendemos con este trabajo es una caracterización de la Comunidad Autónoma de Castilla y León según la producción agrícola (cereales, legumbres, remolacha, patata y hortalizas) y la producción cárnica (bovino, ovino, porcino y caprina) durante el período 1985-1993.

Para dicho propósito se ha utilizado la metodología STATIS (Structuration des Tableaux A Trois Indices de la Statistique), técnica que se enmarca dentro del Análisis de Datos Multivariantes. Describiremos la técnica estadística primero para aplicarla después a un serie de datos reales.

2. METODOLOGÍA STATIS

Permite analizar simultáneamente varias matrices de datos. Comenzó a desarrollarse a finales de los años 70 en Francia por el Equipo de Y. Escoufier. Como primeros trabajos destacamos la interestructura expuesta en PAGES-ESCOUFIER-CAZES (1976) y CAILLIEZ-PAGES (1976) basado en los coeficientes de asociación entre tablas propuestas por Y. ESCOUFIER (1970). Desarrollada por H. L'Hermier des Plantes en su tesis en 1976, el método STATIS ha sido comparado con otras técnicas propuestas por CARROLL-CHANG (1970), TUCKER (1966) y GOWER (1975) y por F. GLACON en su tesis en 1981.

2.1. Descripción

Partimos de una serie de datos sobre un conjunto de N observaciones caracterizadas por p variables durante T períodos.

Supondremos que las matrices de datos de cada período están construidas con variables centradas y reducidas.

Sean $Z_1, Z_2 \dots Z_T$ matrices de datos de orden $p_1 \times N, p_2 \times N \dots p_T \times N$. Cada una de estas matrices corresponde a un período objeto de estudio y caracteriza a N observaciones (las mismas en cada período) por un conjunto de variables que pueden variar de un período a otro.

Para caracterizar los períodos t con matrices del mismo orden, construiremos matrices cuadradas de orden N, S_t , siendo:

$$S_t = \frac{1}{N} Z_t^T Z_t$$

2.2. Ejes factoriales para representar los períodos

Las matrices S_t se pueden transformar en vectores columnas S_t^* de N^2 elementos:

$$\frac{1}{N} (Z_1^T) Z_1^1, \frac{1}{N} (Z_2^T) Z_2^2 \dots \frac{1}{N} (Z_t^T) Z_t^t \dots \frac{1}{N} (Z_N^T) Z_t^N$$

donde Z_t^i representa el vector columna cuyas componentes son las coordenadas del individuo i en las variables que lo caracterizan en el período t .

Asociando a cada período t un vector S_t^* , los distintos períodos pueden considerarse como puntos en el espacio de N^2 dimensiones, y la matriz de datos:

$$S_{N^2 \times T} = [S_1^*, S_2^* \dots S_t^* \dots S_T^*]$$

puede ser el punto de partida para obtener los ejes factoriales de la nube de puntos-períodos $L = \{1, 2 \dots t \dots T\}$.

Siguiendo los mismos principios que en el análisis de componentes principales para cuadros de doble entrada, dando a cada período un peso unidad, dichos ejes serán los vectores propios de la matriz de inercia S_L :

$$S_L = S^T S$$

Llamando a_k ($k = 1, 2 \dots T$) a los vectores propios unitarios de S_L , asociados a los valores propios $\lambda_k^{(a)}$, las coordenadas de los períodos en los ejes factoriales serán:

$$G_k = \begin{bmatrix} G_{k1} \\ G_{k2} \\ \vdots \\ G_{kT} \end{bmatrix} = \sqrt{\lambda_k^{(a)}} a_k$$

Para concluir con los resultados relativos a los períodos, que en la técnica STATIS se denominan la "interestructura", señalaremos que puede construirse la nube de puntos-períodos para que como las variables en el análisis clásico de componentes principales se sitúen en la periferia de una esfera de radio unidad. Para ello basta transformar los vectores

S_t^* para que la norma sea 1. Es decir, en vez de caracterizar los periodos mediante los vectores S_t^* , hacerlo mediante los vectores S_t^u :

$$S_t^u = \frac{1}{\sqrt{(S_t^*)' S_t^*}} S_t^*$$

de este modo, $d^2(t,0) = 1$

Los elementos de la matriz de inercia correspondiente a esta caracterización de los periodos serán:

$$(S_t^u)' S_t^u = \frac{\text{Traza } (S_t S_t')}{\sqrt{\text{Traza } (S_t^2) \text{ Traza } (S_t'^2)}}$$

A estas magnitudes se les suele representar mediante lo que se denomina coeficientes RV :

$$RV(S_t, S_t') = (S_t^u)' S_t^u$$

Los términos de la diagonal principal de esta nueva matriz de inercia serán iguales a 1.

2.3. La matriz compromiso

Se denomina matriz compromiso o simplemente compromiso a la matriz cuadrada C de orden N siguiente:

$$C = \sum_{t=1}^T a_{1t} S_t$$

siendo a_{1t} los elementos del primer vector propio de la matriz S_t . Esta matriz C constituye el mejor resumen de las matrices S_t .

Por otra parte, C es una matriz definida no negativa, ya que es una combinación lineal de matrices definidas no negativas en la que los coeficientes a_{1t} son positivos.

Más clara resulta la interpretación de C si se observa que es la matriz de inercia asociada a la matriz de datos siguientes:

$$\frac{1}{\sqrt{N}} Z' = \frac{1}{N} \left[\sqrt{a_{11}} Z'_1 \sqrt{a_{12}} Z'_2 \dots \sqrt{a_{1T}} Z'_T \right]$$

Esta matriz no es más que una yuxtaposición de las matrices de datos de cada periodo t , multiplicadas por $\sqrt{a_{1t}}$, y verifica la igualdad:

$$C = \frac{1}{N} Z' Z$$

2.4. Representación de las variables

Vista la analogía de la matriz compromiso con la matriz de inercia de la nube de puntos variables en un análisis de componentes principales, llamando a los vectores propios unitarios de la matriz C asociadas a valores propios λ_k , V_k ($k=1, 2 \dots N$), las coordenadas de las variables en los distintos periodos se obtendrán mediante la expresión:

$$G_k = \frac{1}{\sqrt{N}} Z V_k$$

siendo G_k un vector de $\sum_{t=1}^T p_t$ componentes:

$$G_{k_1}^{(1)}, G_{k_2}^{(1)} \dots G_{k_{p_1}}^{(1)} \dots G_{k_1}^{(2)}, G_{k_2}^{(2)} \dots G_{k_{p_2}}^{(2)} \dots G_{k_1}^{(T)}, G_{k_2}^{(T)} \dots G_{k_{p_T}}^{(T)}$$

donde $G_{k_j}^{(t)}$ representa la coordenada de la variable j del período t en el eje k .

2.5. Representación de los puntos compromiso

La matriz de datos $\frac{1}{\sqrt{N}} Z'$ representa N observaciones caracterizadas por $\sum p_t$ variables.

Las proyecciones de estas N observaciones en el eje k serán:

$$F_k = \begin{bmatrix} F_{k1} \\ F_{k2} \\ \vdots \\ F_{kn} \end{bmatrix} = \sqrt{N \lambda_k} V_k$$

Estas coordenadas pueden considerarse como proyecciones de las caracterizaciones medias de las N observaciones en los T periodos, y se denominan puntos compromiso.

La interpretación de las componentes F_k puede hacerse a partir de los coeficientes de correlación de las variables $z_j^{(t)}$ con F_k , y que en este caso valen:

$$\Phi_{z_j^{(t)} F_k} = \frac{1}{\sqrt{a_{11}}} G_{kj}^{(t)}$$

2.6. Representación de las trayectorias las observaciones

Los resultados obtenidos hasta ahora no permiten una representación de las trayectorias de las observaciones, en los ejes factoriales, es decir, una representación de las observaciones en cada uno de los periodos objeto de estudio. Una solución a este problema es asociar la proyección de la observación i en el período t sobre el factor k ($F_{ki}^{(t)}$) al i -ésimo elemento del vector $F_k^{(t)}$, obtenido mediante la expresión siguiente:

$$F_k^{(0)} = \begin{bmatrix} F_{k1}^{(0)} \\ F_{k2}^{(0)} \\ \vdots \\ F_{ki}^{(0)} \\ \vdots \\ F_{kN}^{(0)} \end{bmatrix} = \sqrt{\frac{N}{\lambda_k}} S_i V_k$$

De este modo se verifica: $\sum_{t=1}^T a_{it} F_k^{(0)} = \sqrt{\frac{N}{\lambda_k}} \sum_{t=1}^T a_{it} S_t V_k = \sqrt{\frac{N}{\lambda_k}} CV_k$

Ahora bien, como $CV_k = \lambda_k V_k$, obtendremos: $\sum_{t=1}^T a_{it} F_k^{(0)} = \sqrt{N \lambda_k} V_k = F_k$

Vemos pues que en estas condiciones el vector de proyecciones de los puntos-compromiso sobre el eje k puede considerarse como baricentro de los vectores proyecciones de las observaciones en los diversos períodos objeto de estudio (cada uno de ellos ponderado por a_{it}).

3. APLICACIÓN

El estudio trata de ver la evolución de las provincias de Castilla y León respecto a la producción agrícola (cereales, legumbres, patata, remolacha y hortalizas) y a la producción cárnica (bovina, ovina, caprina y porcina) durante el período 1985-1993.

La fuente de información utilizada ha sido el *Anuario de Estadística Agraria de Castilla y León 1993* publicada por la Junta de Castilla y León.

Analizaremos primero la producción agrícola y después la producción cárnica, utilizando en ambos casos la metodología STATIS.

3.1. Agricultura. Producción agrícola

La exposición de los resultados la haremos siguiendo las mismas fases de que consta el método. En primer lugar se analiza la evolución global (inter-estructura), aquí se comparan entre sí las 9 matrices estudiadas correspondientes a los años de período 1985-1993 y se calcula la matriz compromiso, que es una matriz intermedia entre las iniciales. A continuación se estudian las relaciones entre las variables y su evolución, y se interpretan los ejes principales de la matriz compromiso (que constituye lo fundamental en la intra-estructura). Por último, se analiza la posición relativa media de las provincias y las trayectorias de éstas durante el período.

La tabla 1 representa de manera sintética las cinco variables que hemos utilizado y las siglas con que serán identificadas.

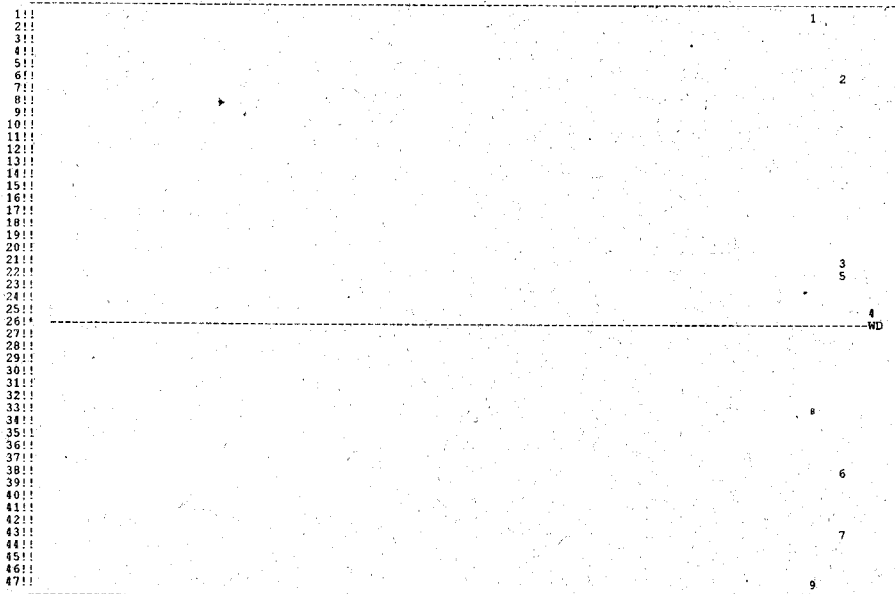
Tabla 1

VARIABLE	SIGLA
Miles de Tm de cereales	CER
Miles de Tm de legumbres	LEG
Miles de Tm de patata	PAT
Miles de Tm de remolacha	REM
Miles de Tm de hortalizas	HORT
PROVINCIA	SIGLA
Ávila	AV
Burgos	BU
León	LE
Palencia	PA
Salamanca	SA
Segovia	SG
Soria	SO
Valladolid	VA
Zamora	ZA

3.1.1. La inter-estructura

La figura 1 proporciona una representación de los años estudiados que explica el 96.14% de la inercia de la matriz de covarianzas entre las matrices.

Figura 1



Los ángulos que forman los vectores que unen los puntos que representan los años con el origen se interpretan en términos de correlación entre las matrices, y las normas de dichos vectores dan una medida global de las correlaciones que se establecen entre las variables del año correspondiente.

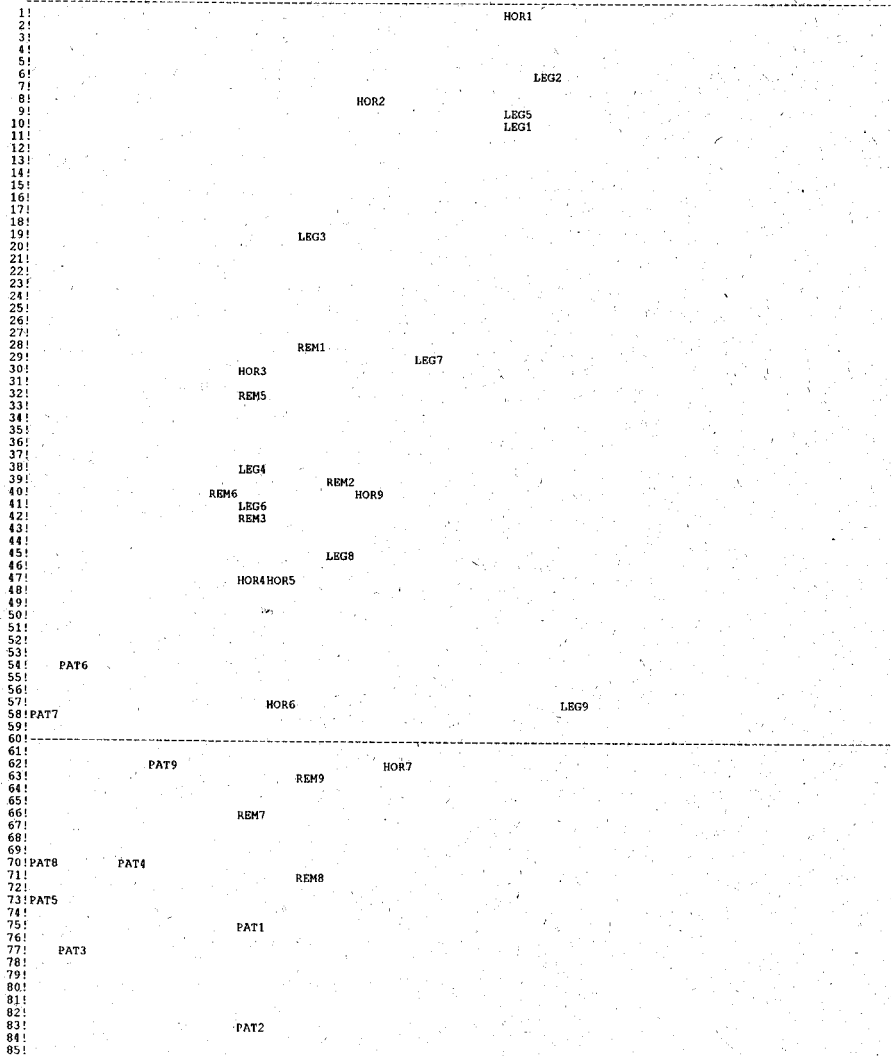
Las matrices tienen estructuras próximas entre sí al hallarse bien correlacionadas. Por tanto, también tendrán estructuras próximas a la del compromiso o matriz intermedia entre

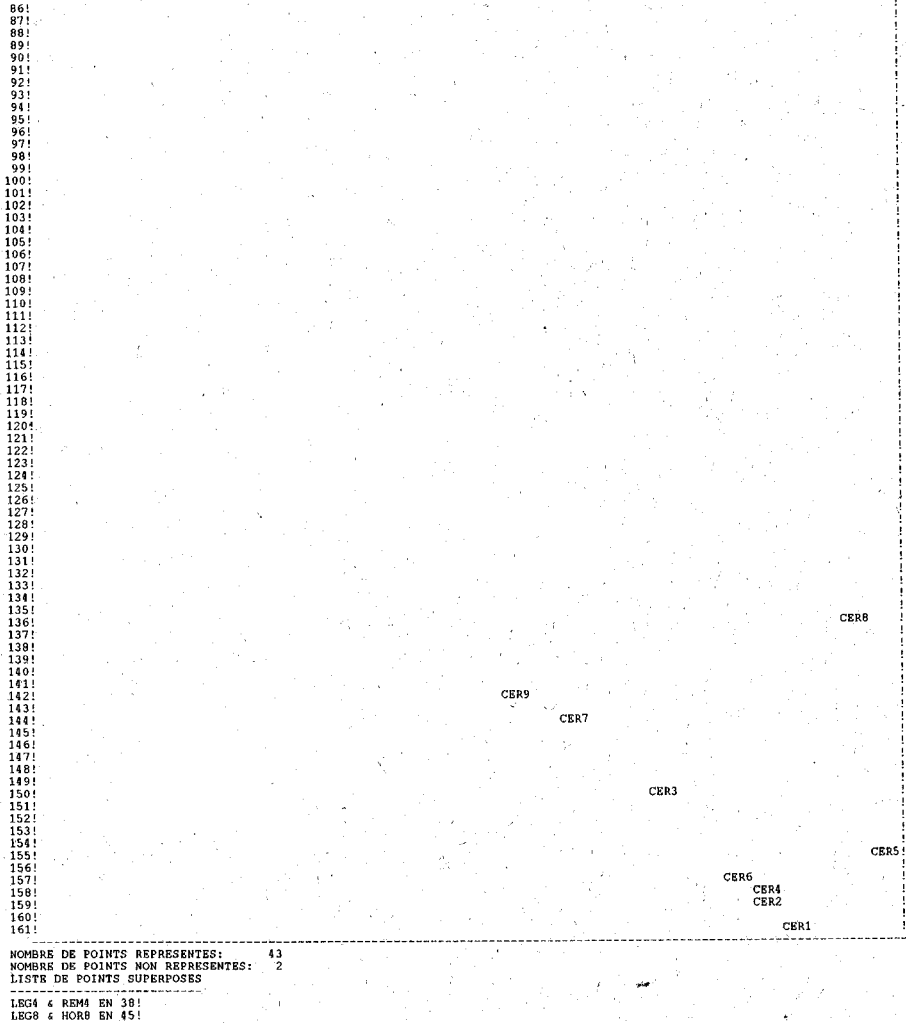
ellas. Esto quiere decir que las posiciones de las variables y de los individuos correspondientes a las diferentes matrices iniciales (años) están, en conjunto, bien representadas.

3.1.2. La intra-estructura

Las principales características de la evolución de las provincias y de las relaciones entre las variables pueden ponerse de manifiesto reteniendo únicamente los dos primeros ejes principales obtenidos del Análisis Factorial en Componentes Principales de la matriz compromiso, que explican el 73.75% de la inercia de dicha matriz.

Figura 2





La figura 2 muestra las proyecciones de las variables en cada uno de los años sobre el plano formado por las dos primeras componentes del compromiso. La interpretación de dicha figura se facilita con la tabla 2 que contiene los coeficientes de correlación de las variables con cada uno de los ejes.

Tabla 2

	CER1	LEG1	PAT1	REM1	HOR1	CER2	LEG2	PAT2	REM2	HOR2
AXE NUMERO 1	-.4114	-.6184	-.8155	-.7733	-.6188	-.4405	-.5964	-.8209	-.7580	-.7321
AXE NUMERO 2	-.3032	-.4523	-.1274	-.2899	-.5307	-.8842	-.4304	-.2034	-.1907	-.4748
0	CER3	LEG3	PAT3	REM3	HOR3	CER4	LEG4	PAT4	REM4	HOR4
AXE NUMERO 1	-.5614	-.7679	-.9456	-.8240	-.8266	-.4378	-.8173	-.9056	-.8272	-.8257
AXE NUMERO 2	-.8022	-.3748	-.1529	-.1615	-.2759	-.8807	-.2000	-.0816	-.2052	-.1221
0	CER5	LEG5	PAT5	REM5	HOR5	CER6	LEG6	PAT6	REM6	HOR6
AXE NUMERO 1	-.3452	-.6176	-.9613	-.8284	-.7965	-.4559	-.8276	-.9591	-.8416	-.8077
AXE NUMERO 2	-.8522	-.4594	-.1128	-.2537	-.1242	-.8645	-.1756	-.0582	-.1861	-.0341
0	CER7	LEG7	PAT7	REM7	HOR7	CER8	LEG8	PAT8	REM8	HOR8
AXE NUMERO 1	-.6202	-.6887	-.9704	-.8201	-.7024	-.3749	-.7568	-.9643	-.7759	-.7491
AXE NUMERO 2	-.7538	-.2805	-.0262	-.0466	-.0120	-.6800	-.1350	-.0846	-.0939	-.1430
0	CER9	LEG9	PAT9	REM9	HOR9					
AXE NUMERO 1	-.6633	-.5769	-.8870	-.7658	-.7261					
AXE NUMERO 2	-.7352	-.0277	-.0170	-.0216	-.1847					

El primer eje explica el 55.51% de la varianza. Dicho eje tiene una fuerte correlación negativa con las variables que representan producción de patata (PAT), remolacha (REM), legumbres (LEG) y hortalizas (HOR). Además, los coeficientes de correlación de las dos últimas variables con el primer eje principal aumentan de los primeros años a los últimos.

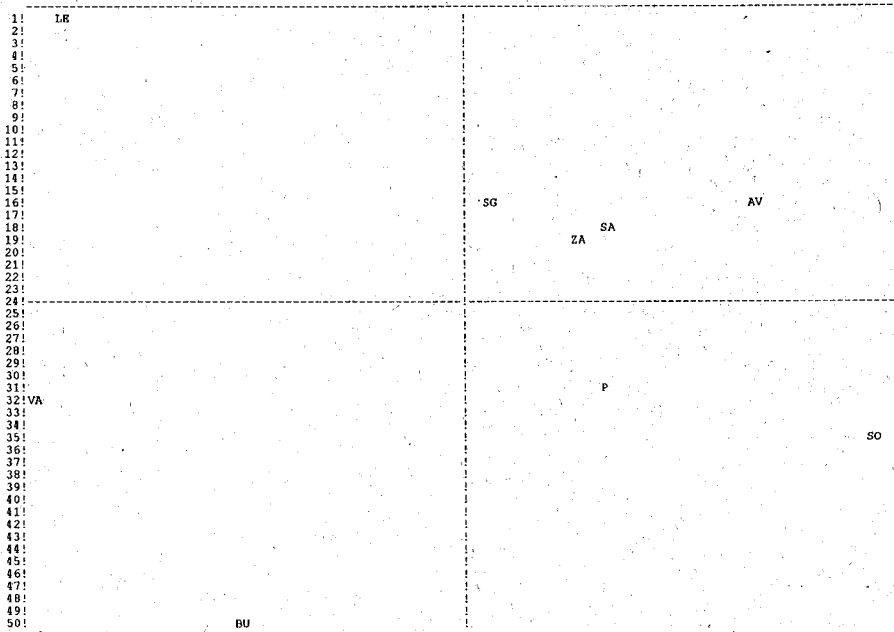
El segundo eje principal, que explica el 18.24% de la varianza del compromiso, tiene una fuerte correlación negativa con la variable que representa la producción de cereales.

3.1.3. Situación relativa y trayectorias de las provincias

En este apartado se estudia la posición de las provincias sobre los dos primeros ejes principales de la matriz compromiso. Se estudia primero, la posición relativa media de las provincias durante el período con los puntos compromiso y, a continuación, se analizan algunas de las trayectorias.

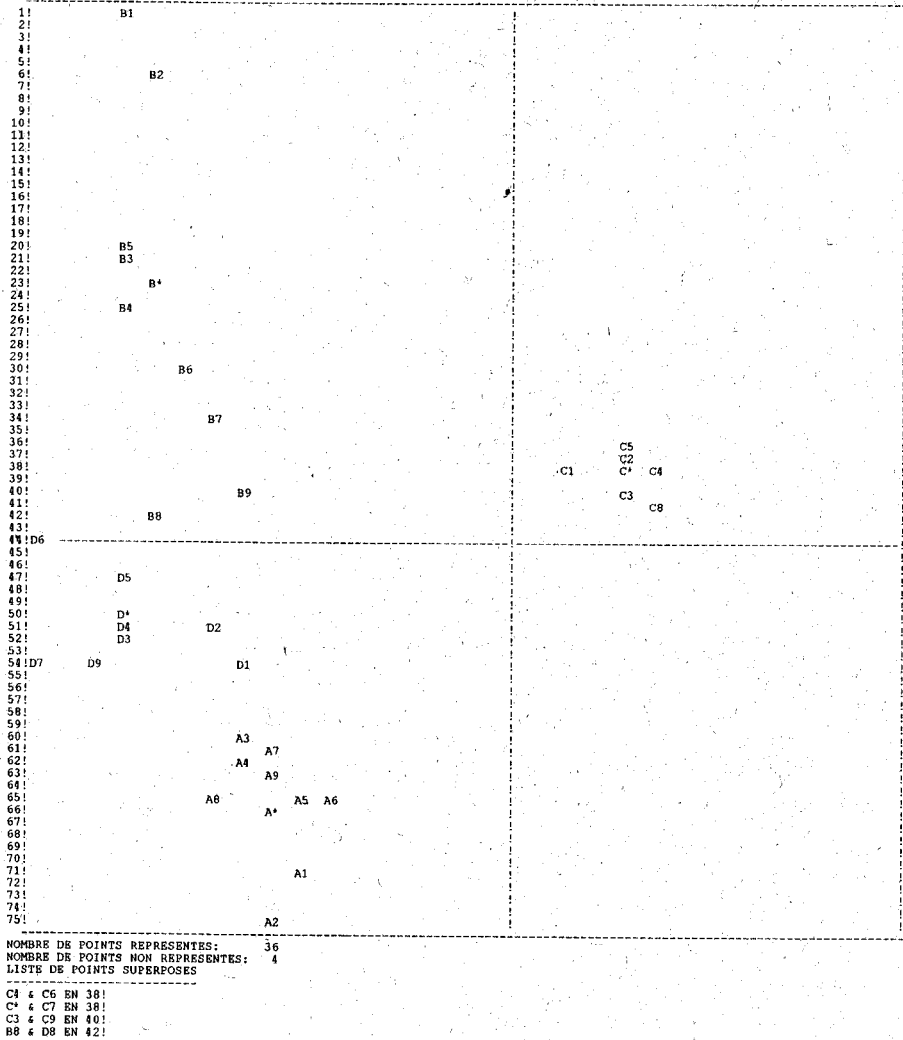
En la figura 3 se representan los puntos compromiso y su interpretación se basa en la que se ha hecho de los ejes.

Figura 3



La figura 4 recoge las evoluciones de León, Valladolid y Burgos sobre el plano formado por los ejes 1 y 2 del compromiso.

Figura 4



3.2. Ganadería. Producción cárnica

La exposición de los resultados se hará de una forma análoga al caso anterior. La tabla 3 representa las cuatro variables que hemos utilizado y las siglas con que son identificadas.

Tabla 3

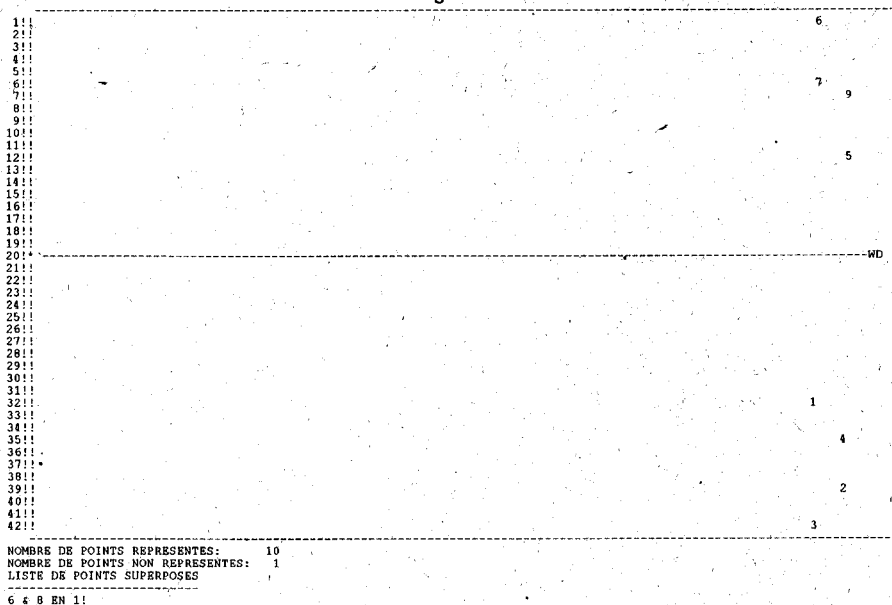
VARIABLE	SIGLA
Tm canal bovino	BOV
Tm canal ovino	OVI
Tm canal caprino	CAP
Tm canal porcino	POR

PROVINCIA	SIGLA
Ávila	AV
Burgos	BU
León	LE
Palencia	PA
Salamanca	SA
Segovia	SG
Soria	SO
Valladolid	VA
Zamora	ZA

3.2.1. La inter-estructura

La figura 5 proporciona una representación de los años estudiados que explica el 95.86% de la inercia de la matriz de covarianzas entre las matrices.

Figura 5

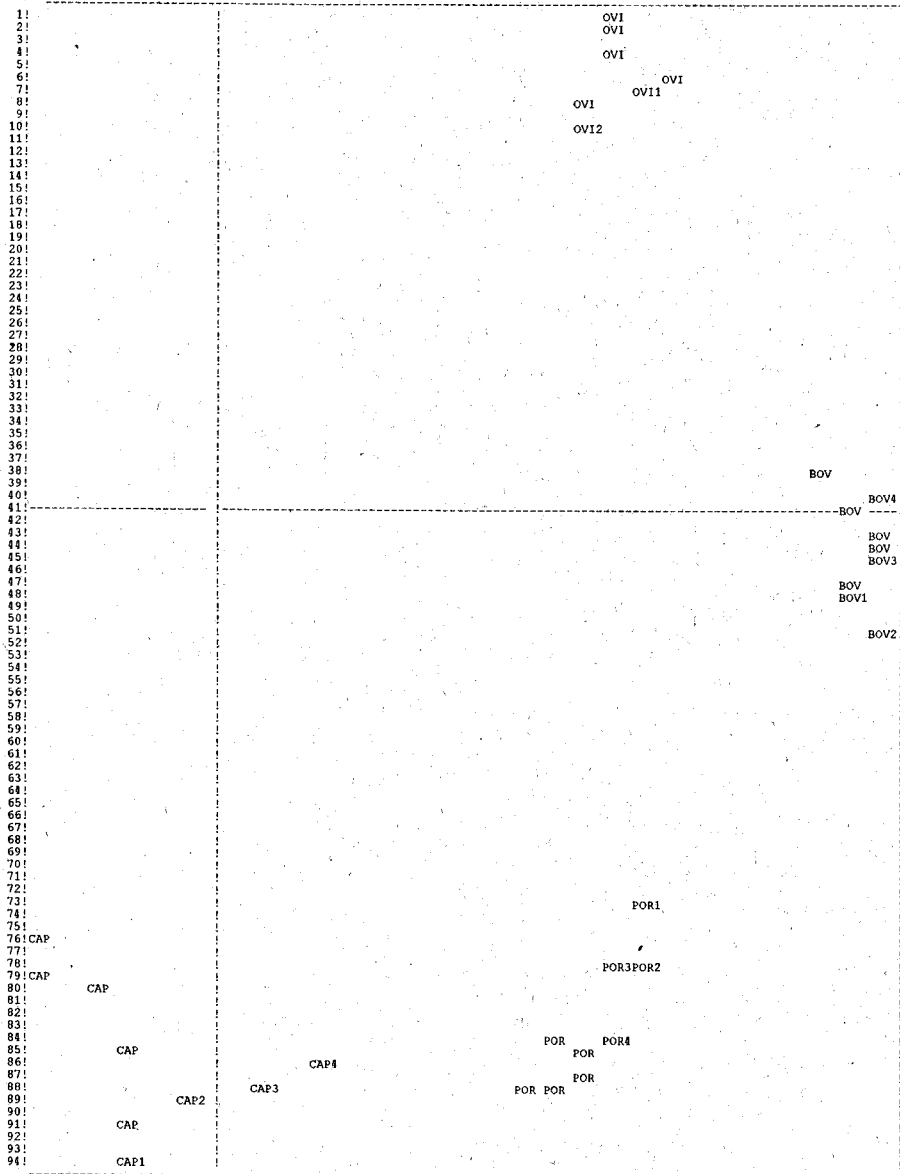


En este caso, las matrices siguen teniendo estructuras próximas entre sí al hallarse bien correlacionadas.

3.2.2. La intra-estructura

Los dos primeros ejes principales obtenidos del Análisis Factorial en Componentes Principales de la matriz compromiso, explican el 77.31% de la inercia de dicha matriz.

Figura 6



NOMBRE DE POINTS REPRESENTES: 34
 NOMBRE DE POINTS NON REPRESENTES: 2
 LISTE DE POINTS SUPERPOSES:

 OVI1 & OVI4 EN 7!
 OVI2 & OVI3 EN 10!

La figura 6 muestra las proyecciones de las variables en cada uno de los años sobre el plano formado por las dos primeras componentes del compromiso. La tabla 4 contiene los coeficientes de correlación de las variables con cada uno de los ejes, lo cual facilita la interpretación de la figura 6.

Tabla 4

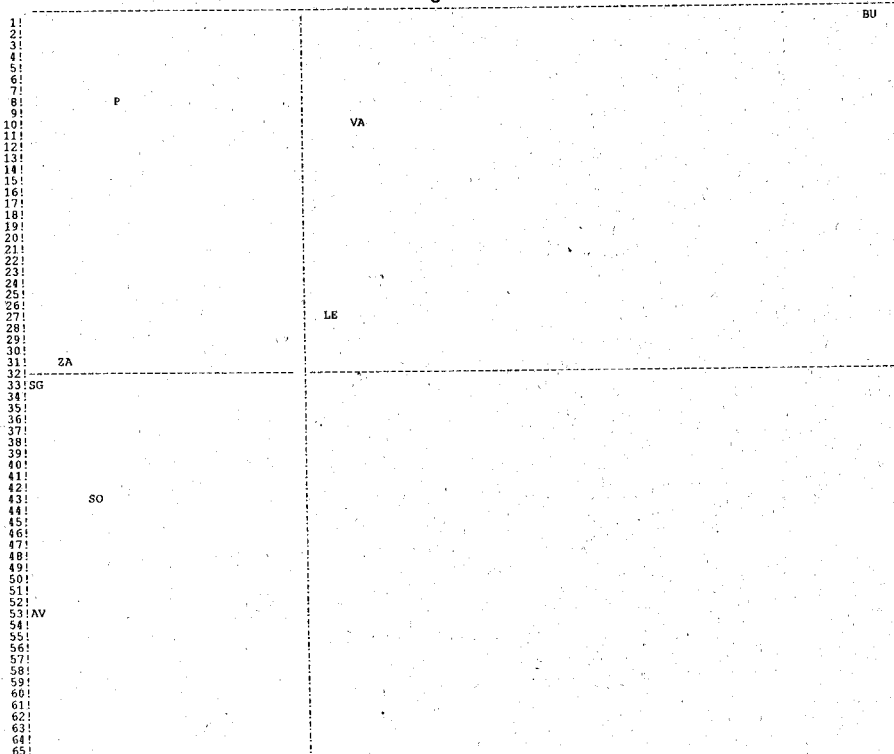
	BOV1	OV11	CAP1	POR1	BOV2	OV12	CAP2	POR2	BOV3	OV13
AXE NUMERO 1	.9265	.6216	-.1414	.6211	.9556	.5627	-.0446	.6082	.9609	.5596
AXE NUMERO 2	-.1165	.5979	-.9193	-.5569	-.1637	.5555	-.8321	-.6458	-.0633	.5490
	CAP3	POR3	BOV4	OV14	CAP4	POR4	BOV	OV1	CAP	POR
AXE NUMERO 1	.0956	.6033	.9603	.6278	.1676	.5793	.9526	.6620	-.1306	.5215
AXE NUMERO 2	-.8130	-.6455	.0304	.6084	-.7867	-.7442	-.0504	.6298	-.8600	-.7414
	BOV	OV1	CAP	POR	BOV	OV1	CAP	POR	BOV	OV1
AXE NUMERO 1	.9195	.5903	-.1325	.5421	.9014	.5315	-.1814	.4581	.8922	.6018
AXE NUMERO 2	-.1029	.6941	-.7647	-.7535	.0071	.5882	-.6782	-.8207	.0689	.6553
	CAP	POR	BOV	OV1	CAP	POR				
AXE NUMERO 1	-.2509	.4840	.9629	.5975	-.2556	.5223				
AXE NUMERO 2	-.6541	-.8165	-.0315	.7062	-.6066	-.8047				

El primer eje explica el 39.12% de la varianza. Dicho eje tiene una fuerte correlación positiva con la variable que representa la producción cárnica de bovino (BOV). El segundo eje principal, que explica el 38.20% de la varianza del compromiso, tiene una fuerte correlación negativa con las variables CAP (producción cárnica de caprino) y POR (producción cárnica de porcino). La variable OVI (producción cárnica de ovino) en los últimos años está más correlacionada con el segundo eje principal.

3.2.3. Situación relativa y trayectorias de las provincias

En la figura 7 se representan los puntos compromiso y su interpretación se basa en la que se ha hecho de los ejes principales.

Figura 7

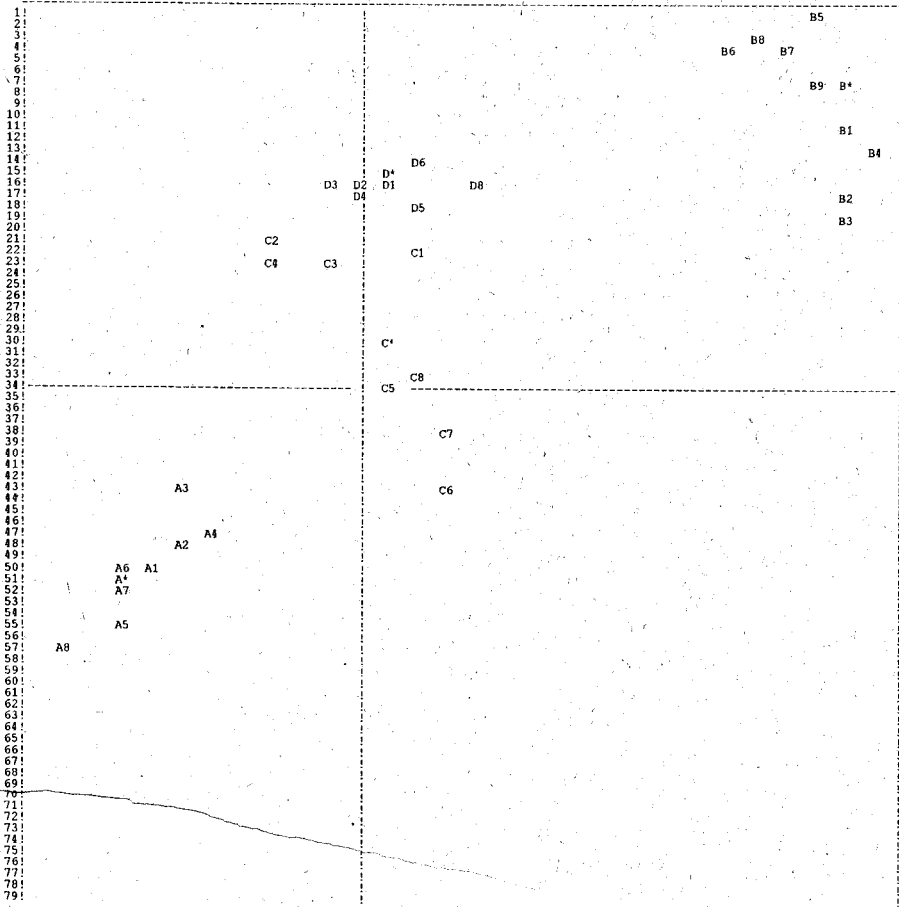


66!
67!
68!
69!
70!
71!
72!
73!
74!
75!
76!
77!
78!
79!
80!
81!
82!
83!
84!

SA

La figura 8 recoge las evoluciones de Salamanca, Avila y Burgos sobre el plano formado por los ejes 1 y 2 del compromiso.

Figura 8



NOMBRE DE PUNTOS REPRESENTES: 36
 NOMBRE DE PUNTOS NON REPRESENTES: 4
 LISTE DE PUNTOS SUPERPOSES

 D6 & D7 EN 14!
 D1 & D9 EN 16!
 C8 & C9 EN 33!
 A* & A9 EN 51!

4. CONCLUSIONES

Con ayuda de las figuras 2 y 3 se comprueba la situación privilegiada de las provincias de León (LE), Burgos (BU) y Valladolid (VA) respecto al resto de las provincias en cuanto a producción agrícola se refiere, destacando Burgos en la producción de cereales (CER), Valladolid en la producción de remolacha (REM) y León en la producción de legumbres (LEG).

De las figuras 6 y 7, deducimos que Salamanca (SA) se caracteriza por la producción cárnica de bovino (BOV) y porcino (POR), Burgos (BU) por la producción cárnica de bovino (BOV) y ovino (OVI), y Avila (AV) por la producción cárnica de caprino (CAP), sobre todo en los últimos años del período estudiado.

De la trayectoria correspondiente a León (figura 4), observamos un aumento en la producción de cereales (CER) a lo largo del período. De la figura 8 deducimos un aumento en la producción cárnica de ovino (OVI) en Burgos, un aumento en la producción cárnica de caprino (CAP) en Avila y un aumento en la producción cárnica de porcino a lo largo del período.

5. BIBLIOGRAFÍA

- BERNARD, M. C., LAVIT, Ch., L'évolution des sociétés rurales du Languedoc-Rousillon (1954-1975): Application de la méthode STATIS, 110e Congrès National des Sociétés Savantes, Géographie, 1985, pp. 183-196.
- JUDEZ ASENSIO, L., Técnicas de Análisis de Datos Multivariantes, Ministerio de Agricultura, Pesca y Alimentación, Madrid, 1989.
- L'HERMIER DES PLANTES, H., Structuration des tableaux à trois indices de la statistique, Thèse de 3^e cycle, Montpellier II, 1976.
- LAVIT, Ch., Analyse Conjointe de Tableaux Quantitatifs, Masson, Paris / Milan / Barcelone / Mexico, 1988.
- NORMAND, B., Analyse de données appliquée au suivi de sujets hémodialyses: Méthode STATIS, Mémoire de DERBM, Paris-Sud, 1987.
- PEREZ HUGALDE, C., Análisis de los resultados económicos de los sectores agrario y no agrario en el proceso de desarrollo, Revista de Estudios Agrò-Sociales, 137 extra, 1986, pp. 317-351.