

# REPRESENTACION GEOMETRICA MULTIDIMENSIONAL DE LA VARIABILIDAD DE CULTIVOS EN LA COMUNIDAD CASTELLANO-LEONESA

S. VICENTE TAVERA\*, M. P. GALINDO\*\*, J. L. VICENTE VILLARDON\*\*\*, A. MARTIN\* , I. BARRERA\*\*, M. J. FERNANDEZ\*\*

\*E. U. de E. Empresariales. \*\* Facultad de Medicina. \*\*\* Facultad de C.C. Económicas y Empresariales  
Departamento de Estadística y Matemática Aplicadas. Universidad de Salamanca.

## INTRODUCCION

Hasta hace poco, el desarrollo de los métodos estadísticos ha estado dominado por los requerimientos de los trabajos con experimentos sobre sistemas altamente controlados donde el error -variaciones no controladas- es muy pequeño.

La mayoría de los procedimientos clásicos están basados explícita o implícitamente en ciertas hipótesis de base acerca de los datos; la validez de los resultados está en relación directa a la validez de los supuestos. Si las condiciones de aplicabilidad se violan, los resultados no son fiables.

La teoría estadística desarrollada para estos patterns de respuesta, posiblemente después de transformaciones, está descrita por los efectos de factores aditivos y una respuesta residual cuya distribución sigue aproximadamente un modelo normal. En los trabajos de economía raramente se pueden suponer estos modelos, ni siquiera cuando los datos se derivan de experimentos controlados.

En las últimas décadas el enfoque de los trabajos teóricos ha sufrido un cambio fantástico. Se han desfasado los modelos estadísticos formales y la teoría de las distribuciones asociadas y han llegado las técnicas descriptivas para inspección de matrices y exploración de patterns en grandes conjuntos de datos. La invasión de los ordenadores trivializa el manejo de las grandes matrices implicadas en los análisis.

Las representaciones gráficas de datos multidimensionales, son probablemente la herramienta de más fácil uso; obras recientes de Análisis Multivariante de reconocido impacto dedican algún capítulo a su estudio; por ejemplo, GNANADESIKAN (1977), EVERITT (1978), BARNETT (1981), JOHNSON and

WICHERN (1982), CHAMBERS et al. (1983). y SEBER (1984).

Es posible hacer una representación multidimensional en el plano siguiendo el procedimiento propuesto por CHERNOFF en 1973 el cual asigna a cada población una cara o rostro humano de forma que la posición, longitud y forma de cada una de los componentes faciales (ojos, cejas, nariz, boca, orejas, pelo, etc) , refleja el comportamiento de una de las variables que intervienen en el estudio. Chernoff aprovecha la capacidad de reconocimiento del observador humano el cual discrimina diariamente sin ningún esfuerzo.

El rango de variabilidad se establece de manera que la estructura global mantenga las características básicas de una cara. La asignación de cada variable a los parámetros faciales es optativa.

La construcción original de Chernoff ha sido desarrollada más tarde y las últimas versiones proponen asignar un diseño a cada una de las dos partes de la cara las cuales se trabajan separadamente (FLURY, 1980); así se explota la sensibilidad del ojo humano para la asimetría.

El análisis de matrices de datos multidimensionales que presentan sus resultados en forma gráfica son potentes herramientas de diagnóstico bien diferentes de los métodos gráficos multivariantes en los cuales sólo se lleva a cabo una transcripción geométrica de los datos.

Complejos procesos algebraicos, asentados sobre métodos numéricos, deben ser definidos antes de pasar del hiperespacio de partida a la bigeometría en la cual se presentan los individuos; sin embargo el resultado final suele ser generalmente un diagrama de dispersión sobre un sistema cartesiano.

En análisis confirmatorios sobre supuestos avalados por la experiencia del investigador o por la literatura científica anterior, el peligro de una interpretación directa de lo que nuestro ojo capta no es tan grave, ya que suele captar lo que sabemos que debe ocurrir; en los análisis exploratorios la interpretación basada en el gráfico resultante es un serio problema que está motivando el encontrarnos con resultados cuya disparidad es injustificable.

La Economía es una Ciencia que se caracteriza esencialmente por analizar fenómenos que están definidos en función de un elevado número de variables, por tanto la importancia de los métodos multivariantes en este campo es evidente. Mientras que en otras Ciencias estos métodos tienen una aplicación ya contrastada, en Economía su uso no se ha generalizado.

De entre las técnicas de representación de datos el Análisis de Componentes principales y el Análisis Factorial de Correspondencias son probablemente las técnicas más populares en este momento; sin embargo, PEARSON descubrió las bases algebraicas sobre las que se asienta estos Análisis en 1901; la primera aplicación hecha por FISHER data de 1940 y en 1941 vuelve a ser propuesta por GUTMAN, aunque ni uno ni otro usa este nombre para referirse al Análisis.

Es fácil encontrar al menos 10 referencias en la literatura de Estadística aplicada en lengua inglesa desde 1955 en las cuales el Análisis de Correspondencias ha sido redescubierto y presentado como algo completamente nuevo y fué en 1973 cuando BENZECRI reinventa la técnica y la bautiza como Análisis de Correspondencias. El estar presentada en lengua francesa y el estilo altamente idiosincrático de las publicaciones de Benzecri no ayudó mucho a la expansión de la técnica por otros países pero esto ha cambiado en buena parte debido a los trabajos de los Psicómetros LINGOES y Mac DONAL y a la obra que GRENAIRE publica en lengua inglesa en 1984.

Los libros de LEBART, MORINEAU, y WARWICH (1984) han sido traducidos al inglés y hay un gran número de "papers" en "Journals" de Estadística con la aproximación Anglo-Americana (usualmente conocida como "log-linear analysis").

Sin duda, la llegada masiva de los ordenadores ha sido decisiva para que el uso de estas técnicas se extienda; programas para efectuar estos Análisis están ya implementados en paquetes de uso común.

En la prestigiosa Facultad de Económicas y Empresariales de la Universidad de Deusto se han defendido recientemente dos tesis doctorales que tienen como eje fundamental la aplicación del Análisis de Componentes Principales y del Análisis Factorial de Correspondencias al estudio de las actitudes (NARVAIZA, 1986) y al estudio socioeconómico de la Rioja (GOMEZ BEZARES, 1989).

Otras técnicas menos conocidas son los Análisis BIPLLOT propuestos por GABRIEL en 1971 y ampliamente desarrollados en sucesivos trabajos GABRIEL 1978, 1990; TSIANCO, 1984). Este procedimiento multivariante proporciona representaciones de  $n$  observaciones (poblaciones) y  $p$  variables en un subespacio de baja dimensión, usualmente el plano o el espacio. Ambas representaciones pueden ser superpuestas para proporcionar información adicional acerca de las relaciones entre variables y observaciones, la cual no está disponible en las representaciones separadas. Este método que puede ser

aplicado a cualquier matriz de datos, a diferencia de lo que ocurre con el análisis Factorial de Correspondencias, ha sido ampliamente probado en otros campos de la Ciencia pero no tenemos referencia de que haya ninguna aplicación en Economía.

GALINDO en 1986 propone otra alternativa de representación simultánea el HJ-BIPLLOT la cual es una extensión de los Biplots de GABRIEL con claras ventajas sobre ellos.

Para un estudio detallado del método consultar GALINDO y CUADRAS 1986 y para tener una información más resumida el trabajo : "Métodos multivariantes de representación simultánea como técnicas de inspección de datos económicos" presentado en este mismo Congreso.

En este trabajo intentamos poner de manifiesto la aplicabilidad de los métodos gráficos multivariantes (Caras de Chernoff ) y del método multivariante gráfico HJ-Biplot a datos económicos :Clasificación de las provincias de la Comunidad de Castilla-León en función a la distribución de la superficie geográfica, según su dedicación.

## DESCRIPCION DE LOS DATOS ANALIZADOS

Los datos han sido tomados del Anuario de Estadística Agraria publicados por el Ministerio de Agricultura Pesca y Alimentación en 1990 y corresponden al censo de 1987.

La superficie de las provincias está dividida en cuatro grupos bien diferenciados:

**-Superficie de cultivo:** comprenden los terrenos ocupados por cultivos temporales, prados temporales para siega o pastoreo y terreno de huerta.

**-Prados y pastizales:** formados por terreno de cubierta herbacea natural continua independientemente si se aprovechan para siega (prado ) o no (pastizal ).

**-Terreno forestal:** formado por monte maderable, monte abierto (dehesas de pasto y arbolado con encinas y otros árboles ) y monte leñoso.

**-Otras superficies:** comprende el erial a pastos, espartizal, terrenos improductivos, superficies no agrarias y rios y lagos.

( A esta clasificación nos referiremos posteriormente en el texto como Clasificación 1).

El terreno de cultivo está a su vez subclasificado en herbaceas, leñosas y barbecho; así, hemos considerado su distribución en porcentaje en relación al total dedicado a cultivo.

El porcentaje correspondiente a prados se ha calculado sobre la superficie total dedicada a prados y pastizales y análogamente para los pastizales.

Para el terreno forestal trabajamos con la distribución de porcentajes en sus tres acepciones: monte maderable, monte abierto y monte leñoso.

"Otras superficies" se subclasifica de manera análoga.

(La definición exacta de estas variables puede encontrarse en el citado Anuario.)

(Esta subdivisión será referida en el texto como Clasificación 2)

Teniendo en cuenta que la mayor parte de las tierras están dedicadas al cultivo, en una segunda fase del estudio se subdivide éste en secano y regadío y se considera la distribución porcentual de la superficie dedicada a los productos que a continuación se especifican, obteniendo 15 variables para este apartado.

**1: % cereales grano en secano. 2: % cereales grano en regadío.**

Estas dos variables representan los tantos por ciento de terrenos de cultivo dedicados a trigo, cebada, avena, centeno, maíz, etc. tanto en invierno como en regadío.

**3.-% leguminosas grano en secano**

**4.-% leguminosas grano regadío**

Estas variables contienen las superficies en % dedicadas a judías secas, habas secas, lentejas, garbanzos, guisantes secos, veza, algarrobas, yeros y otras leguminosas.

**5.-% de patatas en secano; 6.-% de patatas en regadío**

Representan a las superficies dedicadas al cultivo de la patata tanto en secano como en regadío dedicadas para el consumo humano. En este grupo se han excluido otra serie de tubérculos como la batata, el boniato o la chufa por no tener apenas importancia en la Comunidad.

**7.-% de cultivos industriales en secano**

**8.-% de cultivos industriales en regadío**

Se incluyen dentro de este grupo aquellas superficies dedicadas al cultivo de remolacha azucarera, girasol, cártamo, pimiento para pimentón, tabaco y otros cultivos industriales.

**9.-% de cultivos forrajeros en secano**

**10.-% de cultivos forrajeros en regadío**

Se incluyen todos los terrenos dedicados a los cereales de invierno para forraje, alfalfa, trebol, veza para forraje, maíz forrajero, remolacha para forraje, etc.

**11.-% de hortalizas en secano; 12.-% de hortalizas en regadío**

En este grupo se consideran todas las superficies dedicadas a las hortalizas propiamente dichas incluyendo melones y sandías. Se excluyen del mismo las superficies dedicadas a las patatas de consumo por su importancia individual.

**13.-% de frutales en secano; 14.-% de frutales en regadío**

Estas dos variables incluyen información sobre terrenos dedicados a frutales como el manzano, peral, ciruelos, almendros, etc. Se han excluido los cítricos y la vid ya que los primeros apenas tienen incidencia dentro de la Comunidad y los segundos tienen una mención especial en la siguiente variable.

**15.-viñedos de secano.** Se refiere a todas las superficies dedicadas a este producto independientemente del destino a que se dedique las uvas.

( Esta clasificación será referida en el texto como Clasificación 3)

Hemos de hacer mención a que los porcentajes no suman el 100% , ya que se han eliminado las partes de terreno de cultivo dedicados a ciertos productos cuya incidencia en el cómputo general es ínfimo y que supondrían un "ruido".

## **METODOLOGIA ESTADISTICA**

La similitud entre provincias se pone de manifiesto utilizando técnicas gráficas bi y multidimensionales: diagrama de barras y Caras de Chernoff.

Para obtener una ordenación de las provincias en función de la superficie dedicada según las Clasificaciones señaladas en el apartado anterior utilizamos la técnica de representación simultánea HJ-BIPLLOT (GALINDO, op. ct.) la cual permite representar las provincias y las variables como puntos en un subespacio de dimensión dos en el cual es posible interpretar la proximidad entre provincias en términos de similitud, el ángulo entre variables en términos de covariación y la proximidad de un grupo de puntos -variable a un punto-población en términos de preponderancia. El módulo del vector que determina cada punto-variable y el centro de gravedad de la nube proyectada nos informa de la dispersión de esa variable en el muestreo, cuanto más grande es el modulo, más grande la dispersión.

La aproximación a bajo rango se ha hecho siguiendo a GOLUB y REINSCH, 1970 y el tratamiento informático de las correspondientes matrices utilizadas como INPUT se ha efectuado con el Programa RSDM (GALINDO, VICENTE-VILLARDON, BARRERA, 1990) implementado en Pascal sobre ordenador MACINTOSH.

## RESULTADOS

En el diagrama de barras de la figura 1 aparece la distribución porcentual, en función a la superficie geográfica, de las distintas provincias de la Comunidad Castellano-Leonesa según la Clasificación 1. La superficie de cultivo es preponderante en todas las provincias excepto en Avila, León y Soria.

En Avila existe una proporción muy similar entre terreno de cultivo, prados y pastizales y terreno forestal, siendo el terreno de cultivo el más pequeño de los tres. En León el porcentaje entre terreno de cultivo y prados y pastizales es muy similar, quedando los otros dos grupos por debajo de los anteriores. En Soria, las tierras de cultivo y el terreno forestal representan los porcentajes más elevados.

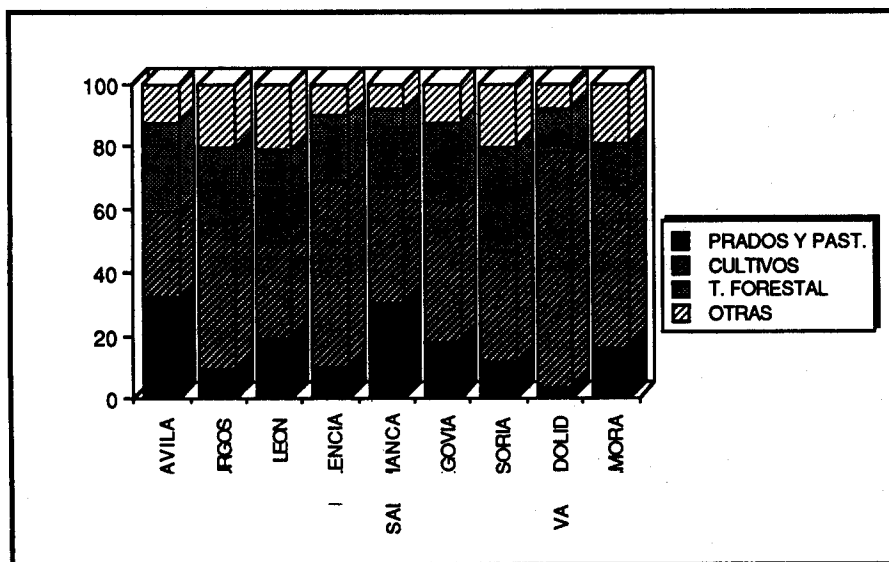


Figura 1: Diagrama de la distribución del suelo en Castilla-León

Tras esta primera visión general analizamos la distribución porcentual correspondiente a la "Clasificación 2" (Ver detalles más arriba).

En este caso se optó por utilizar un método gráfico multidimensional para cada una de las provincias, dado que en esta fase del estudio trabajamos ya con 12 variables. En la figura 2 aparece un pictograma; en cada provincia aparece

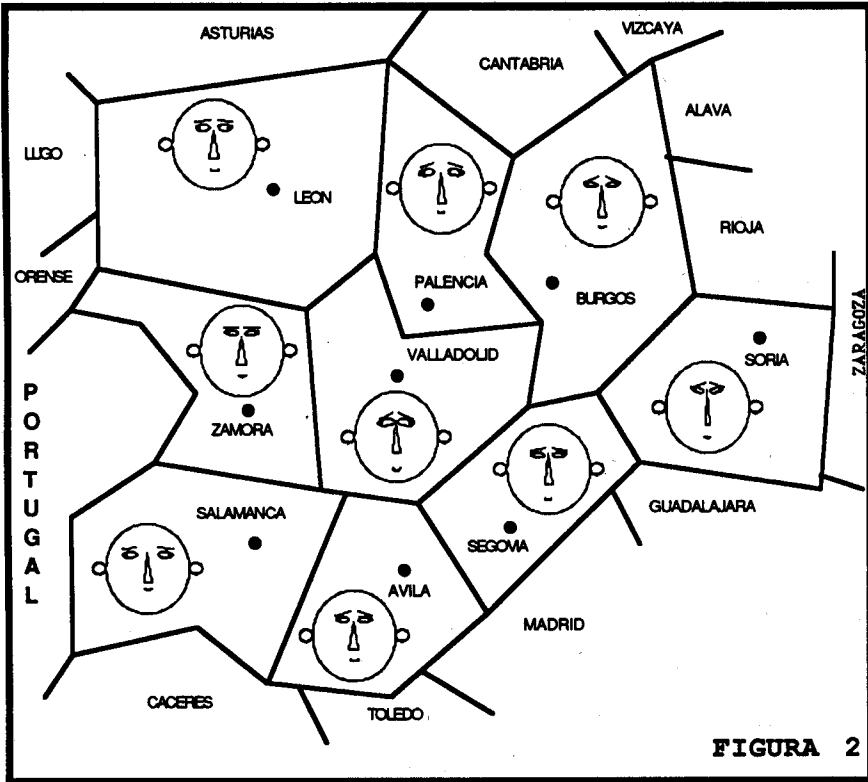


FIGURA 2

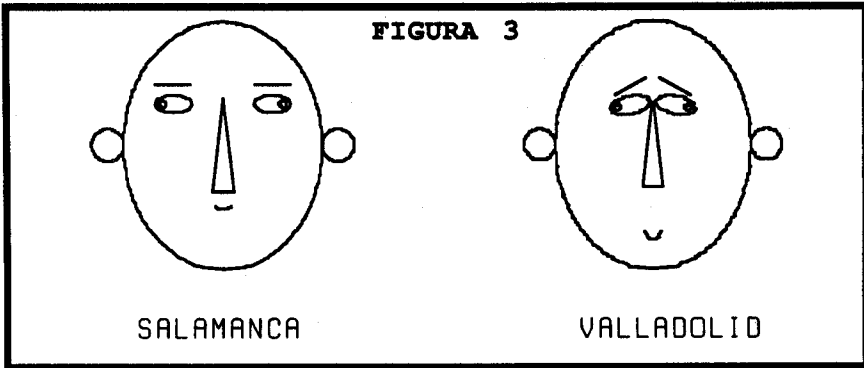


FIGURA 3

SALAMANCA

VALLADOLID

Figura 2: Pictograma de las provincias Castellano-Leonesas utilizando caras de Chernoff  
 Figura 3: Caras de Chernoff correspondientes a las provincias de Salamanca y Valladolid.



una cara de Chernoff que identifica las características de esa provincia.

Se establecieron las siguientes asignaciones de variables a los distintos rasgos de las caras :

VARIABLES	RASGO DE LA CARA
1- Herbáceas	Curvatura de la boca
2- Barbecho	Angulo de la ceja
3- Prados Naturales	Anchura de la nariz
4- Pastizales	Longitud de la nariz
5- Leñosos	Longitud de la boca
6- Monte Maderero	Altura del centro de la boca
7- Monte Abierto	Separación de los ojos
8- Monte Leñoso	Altura del centro de la boca
9- Erial	Inclinación de los ojos
10- Terreno Improductivo	Excentricidad de los ojos
11- Terreno no Agrario	Longitud media de los ojos
12- Ríos y Lagos	Posición de las pupilas

Tabla 1: Asignación de variables a rasgos de la cara y número de identificación de las variables para el análisis multivariante posterior.

Una primera visión del gráfico 2 permite apreciar el gran parecido entre León y Zamora: en el ángulo de las cejas (Var 2 ), curvatura de la boca (Var. 1 ), anchura de la nariz (Var. 3 ) y altura del centro de los ojos (Var. 8 ).

Del mismo modo encontramos parecidos entre Burgos, Soria y Segovia en la inclinación de los ojos ( Var. 9 ), longitud de la nariz (Var. 4) y Curvatura de la boca ( variable 1)

El gráfico 3 nos muestra una vista ampliada de las caras que más diferencias de rasgos presentan; se corresponden con las provincias de Salamanca y Valladolid.

En la figura 4 aparece la proyección en el subespacio de máxima inercia resultado del análisis HJ-Biplot de la matriz correspondiente a la "Clasificación 2"; es decir la matriz que tiene por columnas las 9 provincias de la Comunidad Castilla-León y por filas las 12 variables siguientes : 1(% de herbáceas en supf. de cultivo), 2 (% de barbecho en supf. de cultivo), 3 (% Prados naturales sobre total de prados y pastizales), 4 ( % Pastizales sobre total de prados y pastizales), 5 (% de leñosas en supf. de cultivo), 6 (% Monte maderero en terreno forestal), 7 (% Monte abierto en terreno forestal), 8 ( % Monte leñoso en terreno forestal),9 ( %

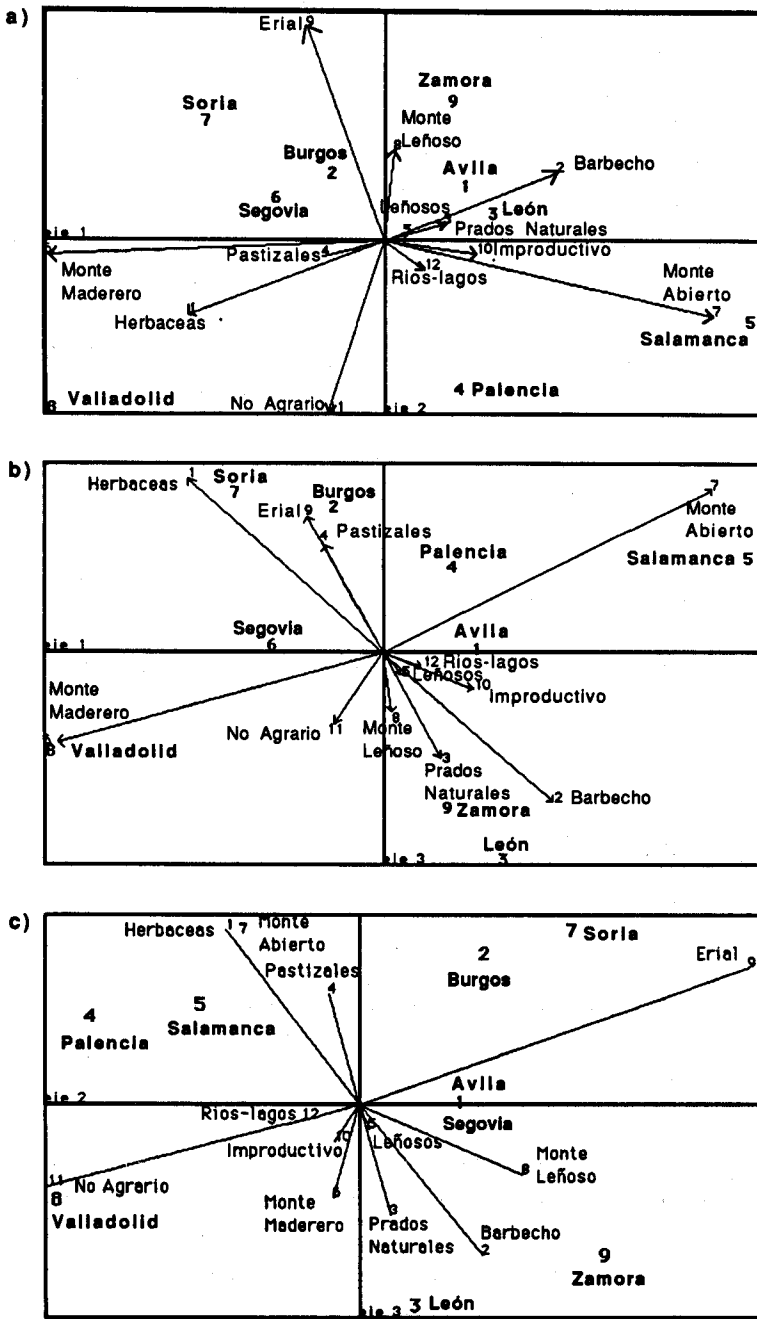


Figura 4 : Diagrama de ordenación de las provincias de la Comunidadde Castilla-León según la clasificación 2 (ver texto)  
a: Proyección en el plano I-II. b: en el plano I-III c: en el plano II-III

Erial sobre "Otras superficies"), 10 ( % Improductivo sobre otras superficies), 11 ( % No agrario sobre otras superficies), 12( % Rios y lagos sobre otras superficies).

En los gráficas (Gráf. 4 a, b, y c) aparecen las variables identificadas con los números antes señalados y una abreviatura de los nombres.

Los tres primeros ejes absorben el 89.47% de la inercia; el primer eje aporta un 47.17%, el segundo un 27.02% y el tercero un 15.29%.

El eje I está prácticamente determinado por "terreno forestal" y "superficie de cultivo" ya que las variables 6 y 7 (Monte Maderero y Monte Abierto) son los que más contribución relativa del factor al elemento (C.R.F.E.) tienen; para el Monte Leñoso, la contribución es más baja. Algo similar al grupo del Terrero Forestal, ocurre con el Terreno de Cultivo, donde las variables 1 y 2 (Herbáceos y Barbecho) son la 3ª y 4ª variables que más C. R. F. E. tienen, mientras que la 3ª variable de este grupo Leñoso (variable 5), su contribución es menos importante. Ver tabla 1

PROVINCIAS							
	EJE1	EJE2	EJE3		EJE1	EJE2	EJE3
1:	260	190	0	2:	93	200	390
3:	230	19	680	4:	76	730	76
5:	830	94	35	6:	310	97	0
7:	410	290	250	8:	640	310	42
9:	67	520	320				

VARIABLES							
	EJE1	EJE2	EJE3		EJE1	EJE2	EJE3
1:	510	130	290	2:	480	130	260
3:	200	28	460	4:	200	28	460
5:	240	31	230	6:	880	3	47
7:	770	74	130	8:	3	380	89
9:	57	800	130	10:	480	19	57
11:	49	830	75	12:	450	430	41

Tabla 1: Contribuciones relativas del Factor al Elemento (C.R.F.E.) (Clasificación 2).

El eje 2 queda determinado por el grupo de "Otras Superficies". Especificando, vemos que las variables que más C. R. F. E. tienen, son las 11 y 9 (Terreno Agrario y Erial), quedando en tercer lugar la variable 12 (Rios y Lagos) cuya contribución es de 430 la cual es algo inferior a la que tiene en el eje 1. Ver tabla 1. La cuarta variable, perteneciente a este grupo, variable 10 (Terreno Improductivo), tiene más contribución en el eje 1 que en el 2 y el 3.

El tercer eje, viene determinado por el grupo de Prados y Pastizales donde sus dos variables: 3 y 4 (Prados Naturales y Pastizales), quedan desmarcados del resto de las variables, teniendo idéntica C. R. F. E. (460).

Como hemos visto, los tres ejes quedan perfectamente determinados por los distintos grupos de superficies.

Si hacemos el análisis de las provincias en función de los ejes, se ve que Salamanca y Valladolid ocupan posiciones extremas sobre eje 1. La primera por la gran importancia que tiene el Monte Abierto en dicha provincia (69% del terreno forestal) y la segunda por el Monte Maderero y los Herbáceos (75% y 89 % respectivamente).

Las posiciones más extremas sobre eje 2 las toman Palencia y Zamora pero estas provincias reciben influencia también de los otros ejes.

En relación al eje 3, León es la provincia con contribución más alta (680), probablemente por su gran porcentaje en Prados Naturales.

Al hacer el análisis conjunto de provincias y variables, en el plano 1-2 , vemos que Salamanca y Valladolid se diferencian del resto y que la primera aparece próxima a la variable 7 (Monte Abierto) y la segunda está en el lado opuesto siendo las variables 6 y 1 (Monte Maderero y Herbáceos), las que más incidencia tienen en dicha provincia.

El eje 2 del plano, divide claramente a dos grupos de provincias; por un lado están Zamora, Palencia, Avila y León, donde las dos primeras se oponen puesto que Zamora está más influenciada por el Erial (69% de Otras Superficies) mientras que Palencia lo está por el Terreno no Agrario (41%).

El otro grupo de provincias es más próximo y homogéneo (situado en el 4º cuadrante) y lo forman Soria, Burgos y Segovia las cuales tienen un alto porcentaje en Erial y Monte Maderero .

El plano 1-3, de nuevo nos separa perfectamente a Salamanca y Valladolid. El eje 3 divide de nuevo a los dos grupos anteriores: Palencia, Avila, Zamora y León donde esta última es la más alejada al ser del eje 3. Zamora y Palencia, se oponen en ese grupo, siendo Avila la peor representada. El otro grupo, de nuevo es más homogéneo, siendo el Erial, Pastizal y Herbáceos quienes lo determinan.

En relación a la variabilidad global, las variables con mayor dispersión en el estudio son la 7 y 6 (Monte Maderero y Monte Abierto) por una parte, y 9 (Erial) por otra, con una Contribución Relativa a la traza de 210, 190 y 150 respetivamente, quedando las variables 1, 11 y 2 (Herbáceos, No Agrario y

Barbecho) a continuación. (Las contribuciones a la traza están valoradas sobre un máximo de 1000)

Otra cuestión a tener en cuenta, es la perfecta correlación negativa que se aprecia entre las variables 1 y 2 (Herbáceas y Barbecho), por una parte y entre las 3 y 4 (Prados Naturales y Pastizales) por otra. Así mismo, hemos de hacer notar que en el plano 1-2, las variables 1 y 2 son independientes (perpendiculares) de la variable 9 (Erial).

Por otra parte, en ambos planos se ve perfectamente la alta correlación negativa entre el Monte Abierto y el Monte Maderero.

### **Análisis de los distintos tipos de cultivo.**

Una vez analizada la distribución de la superficie geográfica de las provincias de Castilla-León, y visto que la superficie de cultivo es la preponderante en la mayoría de las provincias de la Comunidad y, en el análisis hecho anteriormente, el grupo de Terreno de Cultivo presentaba una fuerte influencia en el eje 1, (el de mayor inercia), hemos preferido establecer dicha clasificación en función de las superficies dedicadas a los diferentes tipos de cultivo en cada provincia. Se ha establecido una distribución, en %, en la tierra de cultivo de cada provincia de la Comunidad, en función a la dedicación a los distintos productos.

El análisis **HJ-Biplot** de la matriz correspondiente a la "Clasificación 3"; es decir, la matriz de 9 columnas (provincias) y 15 filas (productos) puso de manifiesto la existencia de un solo eje significativo ya que la absorción de inercia es del 94,90% para el primer eje factorial. (Para ver detalladamente la definición de estas 15 variables ver el apartado anterior correspondiente a la descripción de los datos.)

La variable 1 (Cereales Grano en seco) presenta una contribución al factor 1 de 920, mientras que la siguiente es la 15 (Viñedos de seco) con contribución 12, ambas valoradas sobre mil.

Todas las provincias de la Comunidad quedan determinadas por el eje 1 excepto Segovia que es la única del eje 2. Este desmarque de Segovia, es como consecuencia de que en la variable 7 (Productos Industriales de seco) toma un valor netamente superior al resto de las provincias.

Como consecuencia de la preponderancia de las variables 1 y 7 (Cereales Grano de seco y Cultivos Industriales de seco), es imprescindible eliminar estas variables de la batería activa.

Por razones de implementación informática no se han podido mantener

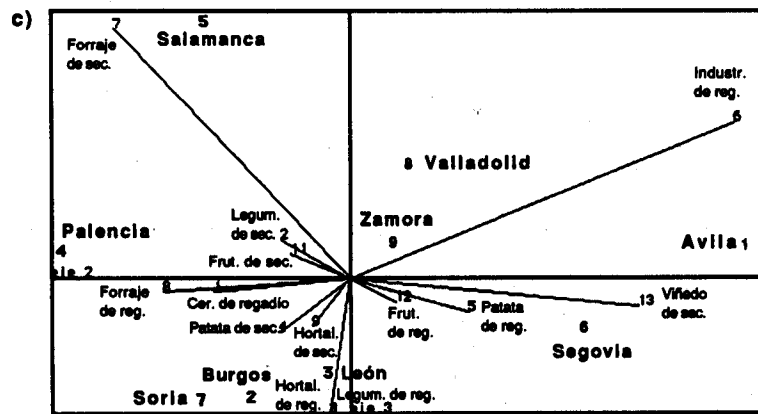
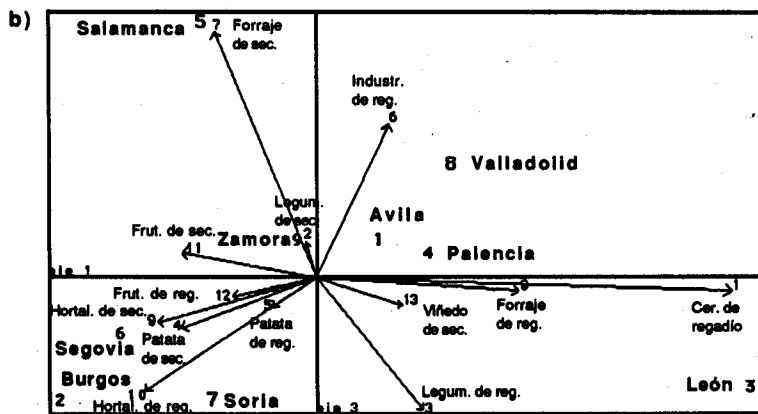
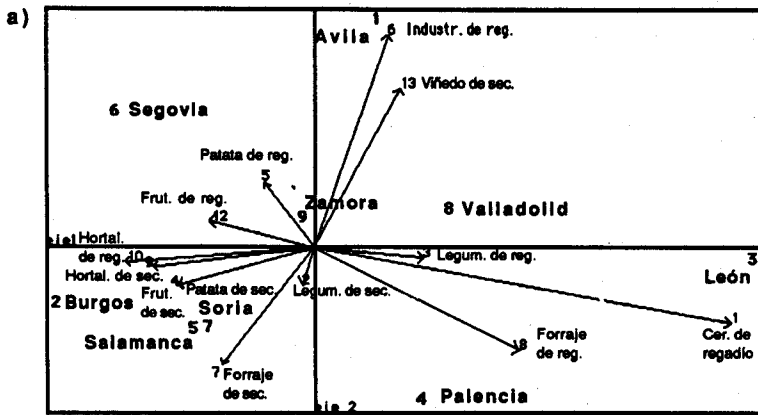


Figura 5 : Diagrama de ordenación de las provincias de la Comunidad de Castilla-León según la clasificación 3 (ver texto)  
 a: Proyección en el plano I-II. b: en el plano I-III c: en el plano II-III

los números del análisis anterior:

Las nuevas equivalencias son

Análisis anterior	Análisis actual
2.....6	1.....5
8.....15	6.....13

Reordenando la matriz y aplicando de nuevo la técnica de representación simultánea **HJ-Biplot** llegamos al siguiente análisis:

La inercia se reparte entre los cuatro primeros ejes, siendo el primero netamente superior al resto. Los cuatro ejes acumulan un 93,03% del total de inercia.

Las variables, cuya C.R. F. E. son más altas en el eje 1, son: 9 Hortalizas de secano, 1 Cereales Grano de regadío, 4 Patatas de secano, 11 Frutales de secano, 8 Forrajeras de regadío y 10 Hortalizas de regadío. Ver tabla 2.

PROVINCIAS							
	EJE1	EJE2	EJE3		EJE1	EJE2	EJE3
1:	52	880	8	2:	770	34	85
3:	920	1	29	4:	250	490	2
5:	240	110	500	6:	600	210	25
7:	360	190	230	8:	380	21	130
9:	26	30	30				

VARIABLES							
	EJE1	EJE2	EJE3		EJE1	EJE2	EJE3
1:	900	26	0	2:	15	160	88
3:	340	3	240	4:	830	55	54
5:	220	310	40	6:	92	660	180
7:	170	250	480	8:	700	160	1
9:	930	12	34	10:	690	4	140
11:	790	44	14	12:	600	33	13
13:	150	440	7				

Tabla 2: Contribuciones relativas del Factor al Elemento (C.R.F.E.) (Clasificación 3).

En el eje 2 destacan dos variables, la 6 y la 13 (Cultivos Industriales de regadío y Viñedos de secano) con C. R. F. E. altas.

En el eje 3, destaca solamente la variable 7 (Forrajeras de secano) con una C. R. F. E. de 480.

La que más destaca en la inercia global, es la variable 1 (Cereales Grano de regadío) con una contribución relativa a la traza de 310, mientras que la segunda más importante es la 6 (Cultivos Industriales de regadío) con una C. R. T. de 99.

El estudio de las columnas (provincias) en este último análisis es más complejo que en los estudios anteriores. León ocupa posición extrema sobre eje 1 en relación a Burgos Soria, Salamanca y Segovia. Ver Gráfica 5a. León destaca por la variable 1 (Cereales Grano de regadío); Burgos por las Hortalizas (tanto de secano como de regadío) y Segovia está menos definida que las anteriores, aunque se asemeja a Burgos. El eje 2, marca las diferencias entre Avila y Palencia. En Avila se dan los valores más altos para las variables 6 (Cultivos Industriales de regadío) y 13 (Viñedos de secano).

El eje 3 separa Salamanca por su importancia en la variable 7 (Forraje de secano) de Soria Burgos y León. Ver gráfica 5 b y c.

Las dos provincias restantes de la Comunidad, Zamora y Valladolid, no quedan bien caracterizadas por la abundancia de ningún cultivo concreto, por lo que podemos inferir que tienen repartida su superficie más equilibradamente que el resto, con respecto a los cultivos mayoritarios de la Comunidad. Hemos de hacer mención a que Valladolid tenía, en el Análisis anterior, una dedicación mayoritaria a los Cereales de secano por lo que su posición en éste puede ser ficticia.

## BIBLIOGRAFIA CONSULTADA

- BARNETT, V. (ed.) (1981). *"Interpreting Multivariate Data"*. Wiley.
- CHAMBERS, J. M. ; CLEVELAND, W. S.; KLEINER, B. y TUKEY, P. A. (1983). *Graphical methods for data analysis*. Wadsworth & Brooks/Cole Publishing Company. Statistics/Probability . Series. Pacific Grove, California..
- CHERNOFF, H.(1973). Using Faces to represent points in k-dimensional space graphically. *J. Amer. Statist. Assoc.*, **68**, 361-368.
- EVERITT, B. (1978). *Graphical Techniques for Multivariate Data*. Heinemann Educational Books, London.
- FLURY, B. (1980). Construction of an asymmetrical facet to represent multivariate data graphically . Technical Report nº 3, Univ. of Berne, Department of Statistics.
- GABRIEL, K. R. (1971). 'The biplot graphic display of matrices with applications to principal component analysis'. *Biometrika*, **58**: 453-467.
- GABRIEL, K. R. (1978). 'Lest squares approximation of matrices by additive and



- multiplicative models'. *J. R. Statist. Soc. Bo.* **40** (2): 186-196..
- GABRIEL, K. R. y ODOROFF, CH. L. (1990). 'Biplots in biomedical research'. *Statistics in Medicine*, Vol. **9**: 469-485..
- GALINDO, M. P. (1986). 'Una alternativa de representación simultánea: HJ-BIPLLOT'. *Questio*. Vol. **10**, Nº **1**: 13-23..
- GALINDO, M. P. y CUADRAS, C. M. (1986). 'Una extensión del método Biplot y su relación con otras técnicas'. *Publicaciones de Bioestadística y Biomatemática. Un. de Barcelona* Nº **17**..
- GNANADESIKAN, R. (1977). *Methods for Statistical Data Analysis of Multivariate Observations*. Wiley, New York.
- GOLUB, G. H. y REINSCH, C. (1970). 'Singular value decomposition and least squares solution'. *Numer. Math.* **14**: 403-420..
- GOMEZ BEZARES, F. (1989). Estudio Unificado del Analisis Multivariante mediante la utilización de proyectores y su aplicación al estudio socio-económico de la Rioja. Tesis Doctoral. Univ. Deusto.
- GRENAIRE, M. J. (1984). *Theory and applications of Correspondence Analysis*. Academic. Press. Inc. London..
- JOHNSON, R.A. AND WICHERN, D.W. (1982). *Applied Multivariate Statistical Analysis*. Prentice-Hall, London.
- LEBART, L.; MORINEAU, A.; and WARWICK, K.M. (1984) *Multivariate Descriptive Statistical Analysis*. Wiley. New York.
- NARVAIZA SOLIS, J.L. (1986). Analisis Multivariante aplicado al estudio de las actitudes. Comparación del análisis por Componentes Principales y por Correspondencias. Tesis Doctoral. Universidad de Deusto.
- SEBER, G.A.F. (1984). *Multivariate Observations*. Wiley.
- TSIANKO, M.C. and GABRIEL, K.R. (1984) Modelling Temperature Data: An illustration of the use of Biplots and Bimodels in Nonlinear Modelling. *Journal of Climate and Applied Meteorology*. **23**, 787-799.