

VALIDACIÓN DE UN MODELO DE TIPIFICACIÓN DE EXPLOTACIONES AGRARIAS MEDIANTE TÉCNICAS DE ANÁLISIS DISCRIMINANTE.

Riveiro, J. A.; Marey, M.F.; Álvarez, C.J.^(p)

Abstract

Galicia (NO Spain), is an example of agriculture characterized by the existence of a variety land uses, different productive models, dimensional diversity and geographical dispersion of farms, in a highly heterogeneous territory in terms of constraints for farming development. Any action aimed to improve production efficiency in this activity, must fulfill the characterization of the production models, and needs of a methodology for data analysis, and the obtention of technical-economical and financial results based on models developed from representative farms in the various areas. For this reason, a methodology of grouping farms in different types was developed, based on variables from the microdata Agricultural Census, and using a combinatory analysis technique subject to constraints. In this paper, is described the validation process of this methodology by applying Discriminant Analysis techniques on population census of commercial Dairy farms in Galicia, using characterizing variables of production systems..

Keywords: Productive Agricultural Management, Rural Development, Discriminant Analysis, Typologies of farms, modelling of farm production systems.

Resumen

Galicia (N.O. de España), es un ejemplo de una agricultura caracterizada por la existencia de multitud de aprovechamientos, diferentes modelos productivos, diversidad dimensional y dispersión territorial de las explotaciones, en un territorio muy heterogéneo en cuanto a aptitudes y limitaciones para el desarrollo de la actividad agraria. Cualquier actuación encaminada a la mejora de la eficiencia productiva en el seno de esta actividad, pasa por la caracterización de los modelos productivos, y necesita disponer de una metodología de análisis de datos y obtención de resultados técnico-económicos y financieros a partir de modelos representativos de las explotaciones existentes en las distintas áreas territoriales. Esta fue la razón que motivó el desarrollo de una metodología para agrupar explotaciones en diferentes tipologías, a partir de las variables que aportan los microdatos de los Censos Agrarios, y utilizando una técnica de análisis combinatorio sometido a restricciones. En el presente trabajo describimos la validación de esta metodología mediante la aplicación de técnicas de Análisis Discriminante, sobre la población censal de explotaciones comerciales de Vacuno Lechero de Galicia, utilizando variables caracterizadoras de los sistemas productivos.

Palabras clave: Ordenación Productiva Agraria, Desarrollo Rural, Análisis Discriminante, Tipologías de explotaciones, Modelización de explotaciones agrarios.

1. Introducción

La agricultura gallega (Noroeste de España), se caracteriza por la existencia de multitud de aprovechamientos y una gran diversidad en la estructura productiva de las explotaciones agrarias. A mediados del siglo XX en Galicia, la actividad agraria se basaba en el

minifundismo y una producción destinada casi íntegramente a la subsistencia, con una gran complejidad estructural y con más del 70% de la población ocupada en este sector, en más de 433.000 explotaciones, [3]. En 1999 se podía contabilizar 270.053 explotaciones en 696.691 ha de superficie agraria útil (2,58 ha. por cada explotación), destinadas en la mayor parte de los casos a la producción para autoconsumo, con dimensiones muy variadas, desde aquellas con menos de 1 ha a las que posibilitaban la comercialización de la producción, [11]. Esta situación regional agraria ha soportado una gran transformación estructural en los últimos años. En la actualidad, la EEEA (Encuesta sobre la Estructura de las Explotaciones Agrícolas) de 2005, [12], muestra que el número de explotaciones se reduce a 91.813 mientras que la superficie de cultivo (SAU) se mantiene en valores similares (en torno a las 700.000 ha.). Por otra parte se conserva una cierta complejidad estructural y normalmente, en cada explotación existe más de una orientación productiva específica, que integra distintos aprovechamientos agrarios, e incluso a veces se confunde el autoconsumo con la producción comercial.

En un escenario tan dinámico en cuanto a transformaciones se refiere, en el año 2002, la Administración de Galicia impulsó la realización de estudios de Ordenación Productiva Agraria, que se concretó en el realizado sobre 21 de las 53 comarcas de Galicia, [22]. Para satisfacer las necesidades derivadas en el desarrollo de los trabajos, surgieron diferentes modelos de ayuda, entre ellos destacó Gestop®, un modelo de análisis de los diferentes aprovechamientos agroforestales existentes y potenciales de Galicia, [19]. Este modelo se utilizó para generar resultados técnico-económicos que se podían integrar con otros indicadores (estructurales, sociales y ambientales), permitiendo desarrollar un sistema de apoyo en la toma de decisiones para priorizar las actividades a potenciar en cada espacio territorial de la región.

La aplicación de Gestop® requería información derivada de la situación actual, para que los resultados que aportaba fueran de utilidad real. Para ello fue necesario desarrollar otros sistemas y modelos para dotar de información, específicamente en lo que se refiere a la tipificación, la clasificación, el análisis territorial y la caracterización de explotaciones agrarias, como herramientas de apoyo en procesos de Ordenación Productiva Agraria.

En la bibliografía internacional se encuentran muchas y diversas propuestas de modelos y sistemas desarrollados como ayuda en la planificación de la actividad agraria. Le Ver & Benoit, [17], idearon un modelo basado en reglas agronómicas cualitativas, que permite describir la organización espacial de usos de la tierra, generando mapas de usos. Hardinan et al., [9], utilizan el análisis Cluster para la identificación y la clasificación de sistemas de explotación en una región de China, lo cual les resultó útil para la determinación de límites geográficos de los diferentes sistemas de explotación, y seleccionar así las áreas apropiadas para la introducción de nuevas tecnologías y también para realizar una taxonomía de sistemas de explotaciones. Köbrich et al., [16], plantean un sistema de taxonomía de explotaciones a partir de situaciones típicas o representativas, mediante el uso técnicas estadísticas de análisis multivariante, seleccionando los factores a considerar en el análisis Cluster. Otra perspectiva de clasificación, la aportan Girard et al., [8], proponiendo un modelo que permite catalogar los sistemas de explotación en base a las prácticas de utilización del suelo, admitiendo el análisis temporal y espacial.

Algunos investigadores consideraron el tamaño de la explotación como una variable adecuada para su clasificación tipológica y el análisis de diferentes aspectos. Tavernier and Tolomeo, [21], analizaron explotaciones según rangos dimensionales, encontrando una correlación positiva significativa entre las pequeñas explotaciones y las prácticas de agricultura sostenible. Demircan et al., [7], realizaron un análisis de la estructura económica de granjas de vacuno en Turquía, basado en su clasificación dimensional. Ello les permitió detectar notables diferencias en los resultados técnicos y económicos obtenidos, en relación con el tamaño de la granja. Judez et al., [15], también se basan en una clasificación

dimensional de las explotaciones a partir del número de UDEs, para analizar los efectos derivados de la aplicación de la llamada “Agenda 2000” en explotaciones de vacuno de carne.

2. Antecedentes.

En un proceso de Ordenación Productiva Agraria es necesaria una identificación y evaluación diferenciada de los sistemas productivos existentes. En los trabajos descritos se clasifican y establecen grupos de explotaciones con alguna finalidad concreta, o bien se estudian sistemas o reglas que permitan dicha clasificación, sin embargo, ninguno de ellos aporta los resultados adecuados a los objetivos que se plantean en el marco de una Ordenación Productiva Agraria. Es por ello que se desarrolló un procedimiento original para el agrupamiento de explotaciones en clases disjuntas, atendiendo a las diferencias existentes entre las explotaciones agrarias de una región (tipificación mediante análisis combinatorio), [20], y su evolución temporal, [1], que partió de la hipótesis de que esta diferenciación va a depender fundamentalmente del número y tipo de aprovechamientos que integran cada explotación (definen su orientación productiva, grado de complejidad y diversificación), y de su dimensión productiva (normalmente relacionada con distintos grados de mecanización, especialización y en definitiva, eficiencia productiva), lo cual dará lugar a los diferentes sistemas productivos.

La metodología de tipificación de explotaciones basada en el Análisis Combinatorio, presenta ventajas respecto de otras técnicas de análisis estadístico. En este sentido, Iraizoz et al., [14], detectaron problemas en la formación de grupos mediante técnicas Cluster, derivados de la multicolinealidad entre variables, lo cual es una característica común de los sistemas agrarios, respecto de las variables que caracterizan las explotaciones. Ello obliga frecuentemente a realizar un análisis previo de componentes principales para eliminar algunas de estas variables. La técnica que queremos validar, basada en el análisis combinatorio, permite la selección de las variables a utilizar conforme a los objetivos de agrupamiento, y establecer las reglas de juego mediante restricciones al proceso de generación de grupos, a partir de la información disponible para cada explotación.

La metodología combinatoria utiliza la principal y más extensa fuente de información en cuanto al volumen de explotaciones considerado, son los Censos Agrarios, siendo el último realizado en España el de 1999, [11], del cual están disponibles los microdatos. El hecho de disponer de una amplia información individualizada para cada explotación agraria, unido a la consiguiente posibilidad de incorporar de un modo relacional, otra información con un grado de agregación igual o mayor, hacen del Censo Agrario la fuente idónea para abordar diferentes estudios o investigaciones y a diferentes escalas.

En los censos, [13], cada explotación viene caracterizada por los siguientes apartados: identificación de la explotación, titularidad, personalidad jurídica y gestión, superficie total y parcelación, régimen de tenencia de la tierra, riego, aprovechamiento de la tierra, cultivos, maquinaria, ganadería, instalaciones para el almacenamiento de abonos, mano de obra y comercialización de la producción. Siendo el Censo Agrario una fuente de referencia para investigadores, ante el volumen de información que aporta y por su carácter internacional, se optó por la utilización de los microdatos, específicamente los del realizado en 1999, [11], como fuente principal de datos para el desarrollo de los procedimientos de tipificación de explotaciones agrarias.

2.1. Tipificación mediante proceso Combinatorio

Se fundamenta en un proceso matemático combinatorio sometido a diferentes restricciones, y se realiza mediante una aplicación desarrollada sobre Excel que genera automáticamente las distintas combinaciones posibles de aprovechamientos, aportando la composición y

número de explotaciones en que está presente cada una de ellas. Este proceso se realiza en sucesivas etapas, con la posibilidad de obtener diferentes resultados intermedios o finales, según se describe por Riveiro et al., [20]. Una vez finalizado, se habrán generado un número determinado de combinaciones de aprovechamientos, cuyo análisis permite establecer las tipologías de explotación representativas, de las cuales son conocidos los aprovechamientos que las integran y el número de explotaciones en que están presentes. Estas tipologías recogen desde combinaciones simples de dos aprovechamientos hasta combinaciones más complejas de varios aprovechamientos con un máximo de 11 (el principal y 10 secundarios).

2.2. Resultados del proceso Combinatorio con vacuno lechero

El Censo Agrario de 1999 recoge para el área de estudio (Galicia, NO España), información de un total de 29.180 explotaciones de vacuno lechero que comercializan leche. De esta población se han excluido todas las explotaciones con menos de 10 unidades productoras, por considerar que no atienden a planteamientos productivos comerciales. De este modo, la población estudiada se reduce a 16.445 explotaciones.

La aplicación de la metodología referenciada permitió diferenciar las explotaciones gallegas con presencia de vacuno lechero, en tres tipologías representativas: (T1) explotaciones de vacuno con una producción lechera vinculada a la tierra y basada en los cultivos forrajeros pratenses, con presencia del cultivo de maíz, independientemente de que existan otras actividades complementarias tales como la recría de vacuno, el cultivo de patatas o aprovechamientos forestales, entre los más corrientes. En estas explotaciones no existirá vacuno de aptitud cárnica en un número superior a 4 animales; (T2) idéntica a la T1 pero sin presencia del cultivo de maíz; y (T3) formada por explotaciones que se diferencian de las anteriores (T1 y T2) por la presencia de vacuno de aptitud cárnica en un número superior a 4 animales.

Por otra parte, en cuanto a la distribución de las explotaciones según las tipologías definidas y dimensiones tomadas, atiende a los valores expresados en la tabla 1. Las explotaciones de las tipologías T1 y T2 suponen conjuntamente el 91% del total de las consideradas. La evolución del sector desde 1999, año al que corresponden los datos del censo, ha llevado a una notable disminución de explotaciones mixtas de la T3, [12], con tendencia a su desaparición, lo cual minimiza la importancia de esta tipología de explotaciones.

El objetivo de este artículo es validar este procedimiento, mediante la utilización de una metodología estadística de Análisis Discriminante (D.A.), que permita contrastar la existencia de diferencias entre los grupos previamente identificados mediante Procesos Combinatorios sometidos a restricciones. El D.A. es una técnica de clasificación multivariante muy recurrida por los investigadores para la clasificación de productos y producciones agrarias con diversos objetivos. Davidson et al., [6], la utilizaron para la comparación del estado ecológico de diferentes modelos agroforestales; Liu et al., [18], lo utilizaron en combinación con técnicas de espectroscopía, para la clasificación de diferentes procedencias de vinos de una misma variedad; Cozzolino et al., [5], para analizar la fiabilidad y exactitud de una técnica de espectroscopía como una herramienta para verificar y autenticar el tipo de ensilaje utilizado en la alimentación de rumiantes; Bowler et al., [4], para estudiar los diferentes modelos de granjas agrícolas del Norte de Inglaterra. El D.A. también fue utilizado para estudiar las diferentes respuestas a enfermedades en relación con distintos tipos de modelos de cultivo, [25] o comparar diferentes modelos de manejo de explotaciones y sus rendimientos económicos, [2]. Por otra parte, Hodge, [10], ya utilizó estas técnicas para estudiar el impacto ambiental que suponían los diferentes tipos de granjas agrícolas, y Somda et al, [23], lo utilizaron para estudiar la viabilidad de las granjas en función de la utilización de la mano de obra familiar y su estructura económica.

3. Metodología de agrupamiento mediante D.A.

La metodología que se describe, fundamentada en el Análisis Discriminante, se utilizará con la finalidad de validar la metodología de tipificación mediante un proceso Combinatorio. Ambas metodologías de tipificación (proceso Combinatorio y D.A.), se realizan a partir de la selección de las explotaciones con un aprovechamiento en común.

El análisis discriminante parte de una tabla de datos de n individuos a los que se les ha medido p variables cuantitativas independientes, que actúan como perfil de características de cada uno de ellos. Una variable cualitativa adicional, dependiente o clasificatoria, con dos ó más categorías, define el grupo al que cada individuo pertenece. Es pues una tabla de $n \cdot (p+1)$ casos en la que cada uno figura con un perfil y una asignación a grupo. A partir la tabla $n \cdot (p+1)$ se obtendrá un modelo matemático discriminante contra el cual será contrastado el perfil del nuevo individuo cuyo grupo se desconoce para, en función de un resultado numérico, ser asignado al grupo más probable.

La función discriminante de Fisher (D), (ecuación 1), se obtiene como función lineal de k variables explicativas. Las puntuaciones discriminantes son los resultados que se obtienen al dar valores a X_1, X_2, \dots, X_k en dicha ecuación. El problema matemático consiste en obtener los coeficientes de ponderación u_j que mejor permitan diferenciar los individuos de los distintos grupos.

$$D = u_1 X_1 + u_2 X_2 + \dots + u_k X_k \quad (1)$$

El punto de corte discriminante (C) (ecuación 2), sirve para clasificar cada individuo en su grupo correspondiente. Los criterios que se consideran para la selección de las variables son el de la Lambda Wilks (λ) (ecuación 3) y el estadístico F. Los resultados se interpretan conforme al siguiente criterio: si el valor de λ se aproxima a la unidad, los grupos estarán mezclados y el conjunto de variables independientes no será adecuado para construir la función discriminante, dado que el mayor porcentaje de variabilidad estaría siendo representado por la variabilidad dentro de los grupos.

$$D = u_1 X_1 + u_2 X_2 + \dots + u_k X_k - C \quad (2)$$

$$\lambda = \frac{\text{Suma Cuadrados intragrupos}}{\text{Suma Cuadrados totales}} \quad (3)$$

El criterio de la Lambda de Wilks identifica las mejores variables discriminantes pero será el estadístico F el que determine que variables se deben considerar en el modelo. Para ello, si $F > 3,84$ la variable se considerará, o de lo contrario será eliminada, El valor del estadístico F se determina mediante la ecuación (4), siendo: n el número de explotaciones; g , el número de grupos; p el número de variables independientes considerado; λ_p la λ calculada antes de incluir la variable seleccionada y λ_{p+1} la λ calculada después de incluir la variable seleccionada.

$$F = \left(\frac{n - g - p}{g - 1} \right) \cdot \left(1 - \frac{\lambda_{p+1}}{\lambda_p} \right) \quad (4)$$

Para una población n con p variables independientes, cada explotación puede ser considerada como un punto en un espacio p -dimensional. Las coordenadas de cada punto se obtendrán a partir de las p variables independientes.

A partir de la representación de las n explotaciones, en un espacio R_p , se trata de obtener un nuevo espacio de menor dimensión (b) tal que, al proyectar la nube de puntos sobre dicho espacio, los puntos correspondientes a explotaciones incluidas un mismo grupo, estén próximos entre sí, y los correspondientes a explotaciones incluidas en distintos grupos, estén alejados. En el caso estudiado solo se dispone de dos grupos (T1 y T2), por lo cual solo se obtendrá una función discriminante (ecuación 5), en la cual, el valor D_i corresponde a la puntuación discriminante para cada explotación, X_p el valor de la variable p en la explotación i , y \hat{a} unos coeficientes estimados de modo que los valores obtenidos para D_i cumplan las siguientes condiciones: a) maximice los autovalores obtenidos mediante la ecuación 6; b) minimice la Lambda de Wilks (λ) calculada mediante la ecuación 3; y c) maximice la correlación canónica de acuerdo con la ecuación 7.

$$D_i = \hat{a}_0 + \hat{a}_1 X_1 + \hat{a}_2 X_2 + \dots + \hat{a}_p X_p \quad (5)$$

$$\text{Autovalores} = \frac{\text{Suma Cuadrados intergrupos}}{\text{Suma Cuadrados intragrupos}} \quad (6)$$

$$n = \sqrt{\frac{\text{Suma Cuadrados intergrupos}}{\text{Suma Cuadrados totales}}} \quad (7)$$

Una vez obtenida la función discriminante (ecuación 5), esta se puede utilizar para efectuar una clasificación de los mismos casos utilizados para obtener la función, permitiendo comprobar así el grado de eficacia de la función para realizar la clasificación. Si los resultados son satisfactorios, la función discriminante se podrá utilizar para clasificar futuros casos (explotaciones) de los que, conociendo los valores de las variables independientes, se desconozca el grupo al que pertenecen.

Cuando se manejan grupos de tamaños diferentes, es necesario compensar la desigualdad a la hora de clasificar los casos. Si se asume que las puntuaciones discriminantes atienden a una distribución normal, se puede calcular la probabilidad asociada a un caso en cada uno de los grupos utilizados en el análisis (probabilidad condicional), ecuación 8. Para clasificar cada nuevo caso del que se desconoce a que grupo pertenece, es necesario comparar las probabilidades condicionales que le corresponden en cada uno de los grupos del análisis, utilizando el teorema de Bayes de la ecuación 9.

$$P = (D > d_i | G = g_k) \quad (8)$$

$$P(g_k | d_i) = \frac{P(d_i | g_k) P(g_k)}{\sum_{k=1}^g P(d_i | g_k) P(g_k)} \quad (9)$$

4. Resultados de validación tipologías mediante D.A.

La validación se efectuará recurriendo a la misma fuente de datos; el censo agrario. La información que este proporciona permite seleccionar algunas variables directas o establecer variables calculadas (como relaciones entre variables directas), que puedan servir para establecer grupos de explotaciones con diferencias relevantes en sus sistemas productivos.

De las tres tipologías definidas a partir de la metodología combinatoria, la tipología T3 no necesita validación sobre las diferencias en su sistema productivo respecto de las otras dos

tipologías, dado que los sistemas de manejo del vacuno de aptitud cárnica son distintos del vacuno lechero y por tanto, su presencia exclusiva en estas explotaciones marcará por sí solo notables diferencias. Por otra parte, dado que además esta tipología está representada por un reducido número de explotaciones, que incluso tienden a desaparecer, se opta por su exclusión y centrar el análisis discriminante en las explotaciones de las tipologías T1 y T2. Estas suponen en su conjunto (tabla 1), un total de 15.004 unidades. De esta cifra un total de 36 explotaciones no son válidas por errores en los valores de alguna de las variables. A partir de las 14.798 explotaciones válidas, se establecen dos grupos: uno con 7.456 explotaciones sobre las que se realiza el análisis discriminante, constituyendo una muestra aleatoria de aproximadamente el 50% del total de explotaciones válidas. Las 7.512 explotaciones restantes se reservan para probar la calidad del modelo obtenido.

Dimensión nº V. L.	Tipología 1		Tipología 2		Tipología 3		Todas	
	Número	%	Número	%	Número	%	Número	%
10-24	3.906	64,1%	6.507	73,0%	1.173	81,4%	11.586	70,5%
25-39	1.462	24,0%	1.723	19,3%	200	13,9%	3.385	20,6%
40-54	462	7,6%	497	5,6%	43	3,0%	1.002	6,1%
55-69	131	2,1%	108	1,2%	7	0,5%	246	1,5%
> 69	133	2,2%	75	0,8%	18	1,2%	226	1,4%
Todas	6.094	100,0%	8.910	100,0%	1.441	100,0%	16.445	100,0%

Tabla 1. Resultados de la tipificación mediante análisis combinatorio.

Variables	T1 (3.041 válidas)		T2 (4.415 válidas)		T3 (7.456 válidas)	
	Media	Desv. típ.	Media	Desv. típ.	Media	Desv. típ.
ST	18,631	13,938	16,416	11,473	17,320	12,583
SAU	13,598	11,123	11,869	8,859	12,574	9,881
UGM_Bov/UTAT	15,056	19,946	16,795	144,411	16,086	111,851
UGM_Bov/SF_Tot	2,562	1,222	3,926	45,272	3,370	34,851
CV_Trac_Pr/SF_Tot	11,813	10,708	18,707	176,607	15,895	136,108
%_SF_Prado	0,433	0,056	0,498	0,009	0,472	0,048
%_Maíz	0,133	0,112	0,004	0,017	0,057	0,096
%_Recría	0,238	0,203	0,202	0,202	0,216	0,203
NºParcelas	27,383	25,215	27,032	23,357	27,175	24,131
NºParcel/ha_ST	1,743	1,494	2,033	1,879	1,914	1,738
Edad_Responsable	45,497	11,127	47,094	11,227	46,442	11,213
S_unitaria_parcel	1,263	3,044	1,139	2,476	1,189	2,722
Sf_Tot/S_Tot	0,710	0,196	0,681	0,223	0,693	0,213

Tabla 2. Estadísticos de grupo.

Inicialmente se seleccionaron las 13 variables siguientes: ST, (Superficie total de la explotación); SAU, (Superficie Agraria Util); UGM_Bov/UTAT, (Unidades de Ganado Mayor Bovino por Unidad de Trabajo Agrícola); UGM_Bov/SF_Tot, (Unidades de Ganado Mayor Bovino por hectárea de Superficie Forrajera Total); CV_Trac_Pr/SF_Tot, (Caballos de Vapor en Tractores Propios por hectárea de Superficie Forrajera Total); %_SF_Prado, (Porcentaje de la Superficie Forrajera ocupada con Praderas); %_Maíz, (Porcentaje de la Superficie Forrajera ocupada con Maíz forrajero); %_Recría, (Porcentaje de animales en fase de Recría); NºParcelas, (Número de Parcelas en la Explotación); NºParcel/ha_ST, (Número de Parcelas por hectárea de Superficie Total de la Explotación);

Edad_Responsable, (Edad del Responsable de la explotación); S_unitaria_parcel, (Superficie unitaria de las parcelas agrícolas) y Sf_Tot/S_Tot, (Relación entre la Superficie forrajera de la explotación y su Superficie Total).

La tabla 2 muestra la distribución y los estadísticos del grupo (T1 y T2) para las 13 variables. La utilización del estadístico F de la ecuación 4, para un valor de $n=7.456$ explotaciones y $g=2$ grupos, permitió seleccionar como variables a incluir en el análisis, las 8 consideradas estadísticamente significativas ($F>3,84$), reflejadas en la tabla 3, que permitirán establecer la pertenencia de las explotaciones a uno u otro grupo analizado, (T1 ó T2). Las demás variables de las 13 inicialmente consideradas han sido excluidas dado que el valor que toma el estadístico F es menor que el valor de salida, ($F<2,71$).

Variable	Tolerancia	F exclusión	(λ)
%_Maíz	0,991	5.735,804	0,978
SAU	0,111	12,925	0,553
Sf_Tot/S_Tot	0,435	66,984	0,558
ST	0,103	19,650	0,554
S_unitaria_parcel	0,778	2,766	0,553
NºParcel/ha_ST	0,398	33,761	0,555
NºParcelas	0,372	26,330	0,554
Edad_Responsable	0,974	5,640	0,553

Tabla 3. Inclusión de variables en el análisis.

La función discriminante indicada en la ecuación 5 se aplica para un valor $i=7.456$ explotaciones y un valor $p=7$ variables independientes. En la tabla 4 se presentan los resultados correspondientes a la prueba de significancia de la función discriminante obtenida (ecuación 5). Los valores de lambda 0,500 son bajos, indicio de diferenciación entre grupos, lo que confirman los valores elevados del estadístico, y sobretodo el valor de la significación inferior al 0,05%.

Paso	(λ)	nºp	nºF	nº Ex.	F exacta		
					Estadístico	G.L.	Significancia
1	0,568	1	1	7.454	5.673,065	7.454	0,000%
2	0,563	2	1	7.454	2.893,854	7.453	0,000%
3	0,560	3	1	7.454	1.949,483	7.452	0,000%
4	0,556	4	1	7.454	1.485,823	7.451	0,000%
5	0,556	5	1	7.454	1.191,413	7.450	0,000%
6	0,555	6	1	7.454	995,553	7.449	0,000%
7	0,553	7	1	7.454	860,213	7.448	0,000%
8	0,553	8	1	7.454	753,861	7.447	0,000%

nºp: número de variables en la función discriminante
nºF: número de funciones discriminantes
nºEx.: número de explotaciones consideradas
G.L.: Grados de libertad

Tabla 4. Prueba de significancia de la función discriminante.

En la tabla 5 se muestran los resultados de todas las comparaciones por pares de grupos, para el estadístico F y su significación. Dado que los valores de F son elevados y el nivel de significación inferior al 0,05 %, se puede indicar que todas las variables incluidas en el análisis discriminante diferencian significativamente las tipologías, dos a dos.

Paso	Grupo	Grupo				G.L.
		T1		T2		
		F	Signif.	F	Signif.	
1	T1			5.673,065	0,000%	7.454
	T2	5.673,065	0,000%			
2	T1			2.893,854	0,000%	7.453
	T2	2.893,854	0,000%			
3	T1			1.949,483	0,000%	7.452
	T2	1.949,483	0,000%			
4	T1			1.485,823	0,000%	7.451
	T2	1.485,823	0,000%			
5	T1			1.191,413	0,000%	7.450
	T2	1.191,413	0,000%			
6	T1			995,553	0,000%	7.449
	T2	995,553	0,000%			
7	T1			860,213	0,000%	7.448
	T2	860,213	0,000%			
8	T1			753,861	0,000%	7.447
	T2	753,861	0,000%			

F: estadístico
Sig.: significancia
G.L.: grados de libertad

Tabla 5. Comparaciones de grupos por pares.

El autovalor obtenido para la función discriminante (ecuación 6) es 0,810, y el valor obtenido para la correlación canónica (ecuación 7) es 0,669. Ambos valores están próximos a la unidad, lo cual es un claro indicio de que la función discriminante obtenida logra diferenciar la media de los grupos T1 y T2, lo cual indica que los puntos correspondientes a las explotaciones de ambos grupos, presentan un buen grado de diferenciación de acuerdo con las puntuaciones discriminantes (Di).

Coef.	Variable	Grupo	
		T1	T2
a1	ST	1,047	1,012
a2	SAU	-1,122	-1,088
a3	%_Maíz	31,385	6,410
a4	NºParcelas	-0,012	-0,023
a5	NºParcel/ha_ST	1,729	1,897
a6	Edad_Responsable	0,423	0,430
a7	S_unitaria_parcel	-0,035	-0,013
a8	Sf_Tot/S_Tot	41,902	40,057
a0	(Constante)	-30,938	-27,760

Tabla 6. Valor de los coeficientes de la función discriminante.

El valor obtenido para la Lambda de Wilks (λ) mediante la ecuación 3, es 0,553, con el que se contrasta de forma jerárquica la significación de la función obtenida mediante la hipótesis nula, obteniendo una significación de 0,00 ($< 0,05$), lo que permite concluir que el modelo discrimina bien entre los dos grupos.

Los centroides sitúan a la tipología T1 en la parte positiva con un valor de la función igual a 1,084, y a la tipología T2 en la parte negativa con un valor -0,747.

En la tabla 6 se muestran los valores de los coeficientes a_i estimados para la ecuación 5 de la función discriminante. Estos permitirán calcular los valores que servirán para determinar en que tipología queda incluida cada explotación de la población analizada. En la tabla 7 se muestran los resultados de clasificación obtenidos para ambos casos, seleccionados y nuevos o no seleccionados. Para los casos seleccionados, se logran clasificar correctamente el 84,7% de las explotaciones. El porcentaje de clasificación correcta de las explotaciones utilizadas en la muestra de validación (nuevos casos) alcanza unos valores del 62,8 % para la tipología T1, y un 99,7% para la tipología T2, lo que supone un 84,8% de las explotaciones. En conjunto se alcanza un buen resultado de clasificación, que se sitúa en el 84,6% de explotaciones cuya clasificación coincide según ambas metodologías; análisis discriminante y proceso combinatorio.

	Grupo	Grupo pronosticado		Total	
		T1	T2		
Casos seleccionados	Recuento	T1	1.918	1.123	3.041
		T2	20	4.413	4.433
	Porcentaje	T1	63,07%	36,93%	100,00%
		T2	0,45%	99,55%	100,00%
Casos no seleccionados	Recuento	T1	1.909	1.133	3.042
		T2	15	4.473	4.488
	Porcentaje	T1	62,75%	37,25%	100,00%
		T2	0,33%	99,67%	100,00%

Tabla 8. Resultados de la clasificación.

Estos resultados son similares e incluso mejores a los obtenidos por otros investigadores mediante técnicas de análisis discriminante, así por ejemplo Liu et al., (2006), consiguieron una tasa global de clasificación correcta de vinos del 76,6 %. Stewart et al, [24], en relación con la predicción de usos del suelo y dependiendo del número de grupos considerados y variables seleccionadas, clasifica correctamente entre el 51,65 y el 62,7 % de los casos.

5. Conclusiones

La metodología desarrollada para agrupar explotaciones según tipologías diferenciadas, basada en un algoritmo combinatorio, queda validada cuando tras la realización del Análisis Discriminante, los resultados muestran una elevada coincidencia en la clasificación.

Esto es de gran relevancia dado que la metodología de agrupamiento validada, basada en un algoritmo combinatorio, permite establecer grupos diferenciados de explotaciones a partir de una fuente pública de información como es el Censo Agrario, abarcando la totalidad de explotaciones agrarias del territorio. Por otra parte, dado que la metodología genera los casos y luego la selección de grupos se realiza mediante un proceso manual de análisis y valoración, las tipologías resultantes son más realistas que las obtenidas mediante otras metodologías de carácter estadístico sometidas a las restricciones de los valores numéricos de las diferentes variables y a la subjetividad de las ponderaciones impuestas por criterios no relacionadas con las características diferenciales de los procesos productivos que rigen las explotaciones.

Referencias

[1] Álvarez C.J., Riveiro J.A., Marey M.F. "Typology, classification and characterization of farms for agricultural production planning". *Spanish Journal of Agricultural Research*. Vol. 6 (1), 2008, pp. 125-136.

- [2] Arriaza M., Gómez-Limón J.A. "Comparative performance of selected mathematical programming models". *Agricultural Systems*. Vol. 77 (2), 2003, pp. 155-171.
- [3] Bouhier A. "*Galicia. Ensaio Xeográfico de Análise e Interpretación dun Vello Complexo Agrario. Tomos I y II*", (Traducción de Benxamin Casal Vila), Xunta de Galicia - Consellería de Agricultura Gandería e Montes, Caixanova. Santiago de Compostela, 2001.
- [4] Bowler I., Gordon C., Alasdair C., Ilbery B., Alastair S. "The development of alternative farm enterprises: A study of family labour farms in the Northern Pennines of England", *Journal of Rural Studies*. Vol. 12 (3), 1996, pp. 285-295.
- [5] Cozzolino D., Fassio A., Restaino E., Fernández E., La Manna A. "Verification of silage type using near-infrared spectroscopy combined with multivariate analysis", *Journal of agricultural and food chemistry*. Vol. 56 (1), 2008, pp. 79-83.
- [6] Davidson N.J., Close D.C., Battaglia M., Churchill K., Ottenschlaeger M., Watson T., Bruce J. "Eucalypt health and agricultural land management within bushland in the Midlands of Tasmania, Australia", *Biological Conservation*. Vol. 139 (3-4), 2007, pp. 439-446.
- [7] Demircan V., Binici T., Koknaroglu H., Aktas A.R. "Economic analysis of different dairy farm sizes in Burdur province in Turkey". *Czech Journal of Animal Science*. Vol. 51 (1), 2006, pp. 8-17.
- [8] Girard N., Bellon S., Hubert B., Lardon S., Moulin C.H., Osty P.L. "Categorising combinations of farmers' land use practices: an approach based on examples of sheep farms in the south of France", *Agronomie*. Vol. 21 (5), 2001, pp. 435-459.
- [9] Hardinan R.T., Lacey R., Yang M.Y. "Use of clusters-analysis for identification and classification of farming systems in Qingyang county, central North China", *Agricultural Systems*. Vol. 33 (2), 1990, pp. 115-125.
- [10] Hodge I. "Application of discriminant-analysis for evaluation of local environmental-impact of livestock production", *Agriculture and Environment*. Vol. 4 (2), 1978, pp. 111-121.
- [11] INE. "*Base de microdatos del Censo Agrario de 1999*", [En línea], En: Internet <<http://www.ine.es>> (Consulta, 28 de octubre de 2005).
- [12] INE-EEEE. "*Encuesta sobre la Estructura de las Explotaciones Agrícolas de 2005*", [En línea], En: Internet <<http://www.ine.es>> (Consulta, 10 de enero de 2007).
- [13] INE-Met. "*Metodología del Censo Agrario*" [En línea], En: Internet <<http://www.ine.es>> (Consulta, 28 de octubre de 2005).
- [14] Iraizoz B., Gorton M., Davidova S. "Segmenting farms for analysing agricultural trajectories: A case study of the Navarra region in Spain", *Agricultural Systems*. Vol. 93 (1-3), 2007, pp. 143-169.
- [15] Judez L., Chaya C., Martinez S., Gonzalez A.A. "Effects of the measures envisaged in "Agenda 2000" on arable crop producers and beef and veal producers: an application of Positive Mathematical Programming to representative farms of a Spanish region". *Agricultural Systems*. Vol. 67 (2), 2001, pp. 121-138.
- [16] Köbrich C., Rehman T., Khan M. "Typification of farming systems for constructing representative farm models: two illustrations of the application of multi-variate analyses in Chile and Pakistan", *Agricultural Systems*. Vol. 76 (1), 2003, pp. 141-157.
- [17] Le Ber F., Benoit M. "Modelling the spatial organization of land use in a farming territory. Example of a village in the Plateau Lorrain". *Agronomie*. Vol. 18 (2), 1998, pp. 103-115.
- [18] Liu L., Cozzolino D., Cynkar W.U., Gishen M., Colby C.B. "Geographic classification of Spanish and Australian tempranillo red wines by visible and near-infrared spectroscopy

combined with multivariate analysis”, *Journal of agricultural and food chemistry*. Vol. 54 (18), 2006, pp. 6754-6759.

[19] Riveiro J.A., Álvarez C.J., Miranda D., Pereira J.M. “Profitability and production requirements for land use allocation of farming and forestry land” *Biosystems Engineering*. Vol. 90 (4), 2005, pp. 477-484.

[20] Riveiro J.A., Álvarez C.J., Marey M.F., Marco J.L., Alvarez C.J. “Procedure for the classification and characterization of farms for agricultural production planning: Application in the Northwest of Spain”, *Computers and Electronics in Agriculture*. Vol. 6 (1), 2008, pp. 169–178.

[21] Tavernier E.A., Tolomeo V. “Farm typology and sustainable agriculture: Does size matter?”, *Journal of sustainable agriculture*. Vol. 24 (2). 2004, pp. 33-46.

[22] UTE EIDO-USC. “*Estudios Comarcales de Ordenación Productiva Agraria de 21 comarcas*”, Xunta de Galicia. Consellería de Política Agroalimentaria e Desenvolvemento Rural. Santiago de Compostela, 2004.

[23] Somda J., Kamuanga M., Tollens E. “Characteristics and economic viability of milk production in the smallholder farming systems in the Gambia”, *Agricultural Systems*. Vol. 85 (1), 2005, pp. 42-58.

[24] Stewart A., Reeds L.G. 1979). “An Application of Discriminant Analysis to Agricultural Land Use Prediction” *Economic Geography*, Vol. 55 (2), pp. 114-122.

[25] Van Schik C., Dijkhuizen A.A., Benedictus G., Barkema H.W., Koole J.L. “Exploratory study on the economic value of a closed farming system on Dutch dairy farms”, *The Veterinary Record*. Vol. 142(10), 1998, pp. 240-252.

Agradecimientos

Los autores agradecen al Ministerio de Educación y Ciencia Español, por la financiación de esta investigación a través del proyecto " Integración de información en un modelo metodológico aplicado a la toma de decisiones en la gestión de la ordenación productiva agraria", con referencia: AGL2006-04789/AGR.

Correspondencia (Para más información contacte con):

José Antonio Riveiro Valiño – GI Proyectos y Planificación
Escuela Politécnica Superior. Universidad de Santiago de Compostela.
Departamento de Ingeniería Agroforestal.
Campus Universitario s/n, 27002, Lugo, Spain.
Phone: +34 982 285900 ext 23262; Fax: +34 982 285926;
E-mail: xosanton@lugo.usc.es