

Identificación de Sistemas Productivos Preponderantes en una zona de la provincia de La Pampa utilizando técnicas de DataMining.

Yanina Bellini Saibene¹, Gabriela Iturrioz¹, Héctor Lorda¹

¹ Estación Experimental Agropecuaria INTA Anguil, Ruta Nac. N° 5, Km 580, (6326) Anguil, La Pampa, Argentina

bellini.yanina@inta.gob.ar, iturrioz.gabriela@inta.gob.ar, lorda.hector@inta.gob.ar

Resumen. Se propuso el enfoque metodológico de descubrimiento de conocimiento en base de datos (KDD) para identificar y describir actividades que integren sistemas productivos. Se realizó clustering sobre siete variables obtenidas del procesamiento de un relevamiento anual a productores agropecuarios de la región noreste de La Pampa, Argentina. Se encontraron tres grupos que representan los sistemas productivos preponderantes en dicha región: “Mixto con Ciclo Completo con Predominancia Invernada” (25% de la superficie), “Ciclo completo con predominancia Cría” (18% de la superficie) y “Ciclo Completo” (47% de la superficie). Se realizaron cálculos de margen bruto para cada grupo. Se concluye que las técnicas de clustering son herramientas válidas para la identificación y caracterización técnica económica de sistemas productivos preponderantes, para zonas agroecológicas homogéneas y que la disponibilidad de un relevamiento anual permitirá replicar la metodología a toda la provincia de La Pampa y analizar su dinámica en el tiempo.

Palabras Claves: Minería de Datos, Margen Bruto, Sistemas Productivos Preponderantes, Clustering.

1 Introducción

En la actividad agropecuaria argentina la administración de los costos de la empresa resulta una estrategia fundamental para hacer frente a las diferentes fuentes de riesgo de la actividad [1]. La variable controlable por el empresario son los costos que la actividad genera ya que, en general, el precio viene dado por la oferta y demanda de los mercados externos.

Conocer y cuantificar estos costos es una condición necesaria para que el productor pueda tomar decisiones tendientes a lograr una mejor asignación de los recursos empleados en pos de una mayor eficiencia y productividad [1]. A nivel de región o sector productivo el análisis de dichos indicadores proporciona información que puede mejorar el proceso de planificación económica a escala regional o provincial [2].

La EEA INTA Anguil genera trimestralmente un informe de los Márgenes Brutos (MB) de los principales productos agropecuarios de la provincia de La Pampa con el objetivo de contribuir con información a la toma de decisiones a diferentes actores del sector agropecuario provincial [2].

El MB es el indicador económico seleccionado frecuentemente para analizar las actividades productivas [3]. Si bien se trata de un análisis parcial por evaluar las actividades en forma independiente y contemplar solo los costos que le son atribuibles [3], es posible estimar el resultado global de la empresa en función de la participación de cada actividad.

Es importante advertir que la limitación de los análisis parciales radica en que las actividades individuales pueden presentar resultados holgados por encima de los costos pero la empresa en su globalidad se puede encontrar en crisis [3].

Los informes realizados para La Pampa presentan el margen bruto y los costos directos de una selección de las actividades/productos más importantes y representativos de cuatro zonas de la provincia [2].

El objetivo de este trabajo es identificar y describir sistemas de producción reales y preponderantes de una zona con condiciones agroecológicas homogéneas en La Pampa utilizando técnicas de Data Mining. El propósito es elaborar indicadores económicos a nivel del sistema en su conjunto e indicadores técnicos productivos de las actividades que componen estos sistemas.

2 Metodología

1.1 Descripción de los Datos

El presente estudio se realizó utilizando datos correspondientes al año 2008 del Registro Provincial Agropecuario (REPAGRO). Esta declaración jurada se realiza anualmente a todos los productores agropecuarios de La Pampa desde la Dirección General de Estadística y Censos, Ministerio de la Producción de La Pampa y está enfocada a relevar estadísticas agropecuarias anuales. Contiene más de 500 variables que indagan sobre aspectos productivos (Tabla 1).

Tabla 1. Resumen de variables incorporadas en el REPAGRO 2008 (Fuente: base de datos y formulario del REPAGRO).

Tema	Cant. Variables
Sistema de producción y actividad principal	2
Existencias, sanidad, mortandad y alimentación según categorías ganaderas (bovinos, caprinos, ovinos, porcinos, equinos)	210
Razas según categorías ganaderas	37
Uso de la tierra (Cultivos de cosecha fina y gruesa, verdeos de invierno y verano, pasturas, monte, campo natural y superficie no utilizable)	84
Sistema de Siembra, fertilización, aplicación herbicidas/fungicidas por cultivo	120

Almacenamiento y comercialización	76
Ubicación catastral y croquis de la explotación	12
Identificación	15
Total	556

El REPAGRO es de libre disponibilidad, ya que anualmente se recogen sus resultados en un informe técnico publicado por la Dirección de Estadística y Censos. Es de cumplimiento obligatorio para el productor agropecuario y de carácter de declaración jurada lo que asegura la disponibilidad de la información en el tiempo y la calidad de la misma. Todas las características anteriores, permiten considerarlo como una herramienta estratégica para el desarrollo y posterior aplicaciones de los modelos resultantes.

1.2 Selección de variables

Del total de variables disponibles, se realizó una preselección con el objetivo de reducir la dimensionalidad y utilizar aquellas que aportan datos útiles al objetivo del trabajo.

Como primer paso se realizó un análisis univariado (utilizando SAS), descartándose 301 variables por presentar el 100% de los casos con el mismo dato (valor cero), las mismas hacen referencia a sanidad animal, alimentación, fertilización, aplicación herbicidas/fungicidas por cultivo, almacenamiento y comercialización y uso de categorías ganaderas.

En un segunda etapa, a partir del análisis de trabajos previos de caracterización, diferenciación y/o definición de sistemas productivos preponderantes, se identificó como variables utilizadas las que se refieren a aspectos productivos (indican qué actividades hace el productor) y de manejo (indican cómo las hace) [2] [4] [5] [6] [7] [8] [9]. Este segundo filtro descarta 68 variables referidas a identificación, ubicación catastral, croquis, razas y mortandad.

Finalmente, del conjunto de variables referidas a datos productivos, se decidió utilizar aquellas que totalizan actividades: totales de cabezas por categoría y totales de hectáreas sembradas por cultivos, los cuales reflejan la intención del productor de realizar esas actividades.

Los filtros de las variables se realizaron utilizando SQL Server 2008 Express por medio de sentencias SQL.

1.3 Selección de casos

En la identificación de sistemas productivos preponderantes los análisis se realizan determinando zonas homogéneas, donde las condiciones agroecológicas son compartidas por todos los sistemas productivos [2] [4] [5] [7] [8]. En este trabajo los datos se agruparon por Microrregiones provinciales y se eligieron las Microrregiones 1 y 3 (figura 1), que constituyen una zona agroecológica de características homogéneas [5] [10]. Las Microrregiones fueron definidas como estrategia central del desarrollo provincial, lo que generará un mayor equilibrio territorial y una mejor

distribución de las posibilidades de desarrollo en la Provincia al permitir la valorización integral de todos los recursos. El Centro Regional La Pampa – San Luís del INTA también ha definido su estrategia institucional a nivel de los Territorios que se definen a partir de las Microregiones provinciales [11].

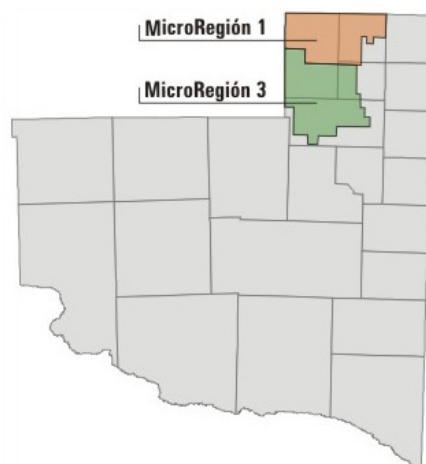


Figura. 1. Microregiones de la provincia La Pampa, Argentina (Ley Provincial N°23.589).

La base de datos de toda la provincia contiene 9.373 casos. Se descartaron aquellos cuya superficie total es menor a 100 ha (por ley provincial, las unidades económicas en la provincia de La Pampa son mayores a 100 ha, predios más pequeños corresponden a quintas cercanas a centros poblados, lugares recreativos, etc. [12]), los que realizan la actividad tambo (fácilmente diferenciables del resto) y los que presentan inconsistencias en los datos (Ej: valor cero en todas las variables de interés, valor cero en el total de la superficie de la explotación agropecuaria, etc.). La cantidad de casos totales final representan el 69% de los productores y el 96% de la superficie relevada de la provincia (La Pampa posee una superficie cercana a las 14.500.000 ha [5]). La población total y la seleccionada para las Microrregiones 1 y 3, quedaron distribuidos según se muestra en la Tabla 2.

Tabla 2. Distribución de productores y superficie para el área de estudio.

Microregiones	Cantidad de Productores	%	Superficie	%
Microregión 1	588	9%	353.511	44%
Microregión 3	594	9%	428.207	5%
La Pampa	6.432	69%	8.783.592	96%

Se volvió a realizar un análisis univariado (utilizando SAS) de las variables de interés y se descartaron 91 que hacen referencia a las categorías ganaderas equinos, porcinos, ovinos y caprinos por tener baja significación relativa en ambas microrregiones [5] y presentar el valor cero en el 90% de los casos en promedio

(Tabla 3), evidencia que respalda la baja significación de las actividades en el área seleccionada para el análisis.

Tabla 3. Porcentaje de casos con valor cero en las variables por Microrregión.

VARIABLES	Microrregion 1	Microrregion 3
Ovinos	93%	88%
Porcinos	93%	90%
Caprinos	99%	96%
Equinos*	77%	77%

**El análisis de los casos disponibles refleja que en dichos casos, la categoría se utiliza como herramienta de trabajo y no como actividad productiva*

El data set final se compone de 38 variables (Tabla 4) y 1.182 casos a procesar. Los filtros sobre los datos se realizaron utilizando SQL Server 2008 Express por medio de sentencias SQL.

Tabla 4. Listado de variables originales utilizadas en el procesamiento.

Tema	
Existencias bovinas (12)	total de cabezas por categoría (terneras menor 1 año, terneros mayor 1 año, vaquillonas 1 a 2 años, vaquillonas mayor 2 años, total vacas, total novillos 1 a 2 años, novillos mayor 2 años, toritos 1 a 2 años, toros mayor 2 años, bueyes y torunos, terneros menor 1 año, total vaquillonas)
Cultivos de cosecha (9)	total de hectáreas por cultivo (trigo, avena para grano, cebada, centeno, girasol, maíz, sorgo granífero, soja, maní)
Cultivos para forraje (15)	total de hectáreas por cultivo (maíz, mijo, sorgo, avena, cebada, centeno, alfalfa pura, alfalfa consociada, otras leguminosas puras, otras leguminosas consociadas, pasto llorón, digitaria, panicum, gramíneas puras, cardenal).
Total de hectáreas (1)	
Número identificador*	

**Esta variable se conserva con el objetivo de identificar a posteriori los casos de cada grupo.*

1.4 Procesamiento

El uso de técnicas de minería de datos para realización de este trabajo se justifica en el hecho que sus herramientas deben funcionar bajo la restricción que los datos disponibles son observacionales (en contraposición a datos experimentales) ya que se trabaja con datos que han sido recolectados para un propósito diferente al análisis de minería de datos a realizar [13].

Para este trabajo se ha utilizado la técnica de clustering que intenta dividir los casos en grupos naturales. Estos grupos presumiblemente reflejan algún patrón que hace que los objetos en un grupo son más similares entre sí (de acuerdo a alguna medida de similitud) y los objetos en grupos separados son menos similares entre sí. [14]

El conocimiento de los grupos puede permitir una descripción sintética de un conjunto de datos multidimensional complejo sustituyendo la descripción de todos los elementos de un grupo por la de un representante característico del mismo. También permite conocer cuáles son las variables importantes que marcan las diferencias entre los grupos y permite generar un modelo que clasifique nuevos casos [14].

El algoritmo utilizado para realizar el análisis es k-means con la implementación del procedimiento FASTCLUS del software SAS. Se estandarizaron las variables con el cálculo: $(\text{variable} - \text{media})/\text{desvío}$. Se realizaron ejecuciones desde 3 a 9 grupos iniciales y diferente orden de los casos.

En las primeras ejecuciones utilizando las 37 variables originales, no se logró separación de grupos, obteniéndose en todos los casos un grupo mayoritario que contenía entre el 87% y 95% de los casos, el cual no lograba separación aún tratándolo individualmente.

Para intentar lograr separación en los datos, en el caso de las variables de actividades de cultivos, las mismas se transformaron de hectáreas a proporciones con respecto al total de superficie de la explotación (tabla 5). Esta transformación permite marcar la diferencia de la decisión sobre el uso del suelo que realiza el productor, sin que influya el tamaño de la explotación en este análisis.

Con respecto a las variables de uso ganadero, las mismas fueron transformadas a proporciones, pero solamente sirvieron para separar grupos cuando se pudo asociarlas a la superficie de la explotación. Para esto se calcularon el índice de cría y la carga animal (tabla 5).

Los filtros de los datos se realizaron utilizando SQL Server 2008 Express por medio de sentencias SQL, las variables originales se procesaron en forma desagregada con estas herramientas lo que permitió generar las 7 variables finales del modelo (Tabla 5).

Tabla 5. Variables finales utilizadas en el clustering.

Variable	Cálculo
Porcentaje de cultivos de cosecha fina	$(\text{Trigo} + \text{Avena} + \text{Cebada} + \text{Centeno}) / \text{Total de Hectáreas}$
Porcentaje de cultivos de cosecha gruesa	$(\text{Girasol} + \text{Maíz} + \text{Sorgo Granífero} + \text{Soja} + \text{Maní}) / \text{Total de Hectáreas}$
Porcentaje de forrajeras anuales	$(\text{Maíz} + \text{Mijo} + \text{Sorgo} + \text{Avena} + \text{Cebada} + \text{Centeno}) / \text{Total de Hectáreas}$
Porcentaje de forrajeras perennes	$(\text{Alfalfa Pura} + \text{Alfalfa Consociada} + \text{Otras Leguminosas Puras} + \text{Otras Leguminosas Consociadas} + \text{Pasto Llorón} + \text{Digitaria} + \text{Panicum} + \text{Gramineas Puras}) / \text{Total de Hectáreas}$
Porcentaje de Caldenal	$\text{Total Caldenal} / \text{Total de Hectáreas}$
Carga animal	$\text{Total cabezas} / \text{Total Superficie Forrajeras}$. Si Total superficie forrajeras es 0 entonces $\text{Total Cabezas} / \text{Total de Hectáreas}$.
Índice de cría	$(\text{Total Vacas} + \text{Total Vaquillonas} + \text{Total Toros} + \text{Total Toritos}) / \text{Total Cabezas}$

Se repitieron las ejecuciones con el procedimiento FASTCLUS (SAS), la estandarización con el cálculo: $(\text{variable} - \text{media})/\text{desvío}$ y la cantidad de grupos: desde 2 a 9 grupos iniciales y diferente orden de los casos.

La denominación “mixto” hace referencia a sistemas que integran actividades agrícolas y ganaderas. Para ser considerado mixto la agricultura deberá contar con una participación en el uso del suelo superior al 30%.

Una vez identificados los sistemas productivos preponderantes (uso del suelo, tamaño de la explotación y dotación ganadera), se realizó el cálculo de MB por actividad participante en el sistema y el MB global de la explotación.

3 Resultados

Con el aumento de la cantidad de grupos aparecen nuevos clusters, los cuales algunos son muy marcados ya que aparecen repetitivamente a pesar de cambiar las semillas y el orden de los casos.

Sobre estas ejecuciones se seleccionó la correspondiente a cinco clusters iniciales de los cuales finalmente fueron seleccionados tres grupos que representan a los sistemas productivos preponderantes de la zona. Para seleccionar los grupos se tuvo en cuenta que representaran más del 15% de la superficie o de los productores de la zona. La figura 2 presenta un gráfico de las componentes principales de los cinco grupos de resultados.

Para evaluar los resultados del agrupamiento, se mide la cohesión (distancia dentro del cluster) y la separación (distancia entre los cluster) [15], con el algoritmo utilizado en este trabajo, minimizar cohesión y maximizar separación son equivalentes por lo que SSE (Sum of Squares for Error: suma de cuadrados del error) es una buena medida del grado de ajuste de los centroides hallados [15]. En cada iteración del algoritmo de las k-medias, se maximiza SSE al calcular los centroides usando la media aritmética; esto no garantiza que los centroides finales obtenidos sean los que minimizan globalmente la función objetivo SSE [15], por lo que una forma de verificar que los grupos realmente se separen entre si es realizar el cálculo de ANOVA [16] [17], verificar con informantes calificados la factibilidad y coherencia de los grupos obtenidos [16] [18] o realizar una muestra de los productores agrupados y realizar una encuesta verificando su pertenencia a cada grupo [18].

El análisis de ANOVA de los grupos obtenidos, presenta que las variables con diferencias significativas entre los grupos son (en orden de importancia): Porcentaje Cosecha Gruesa ($\text{Pr}(> F) = < 2e-16$), Porcentaje Cosecha Fina ($\text{Pr}(> F) = < 2e-16$), Porcentaje Forrajeras Perennes ($\text{Pr}(> F) = < 2e-16$), Porcentaje Forrajera Anual ($\text{Pr}(> F) = 2.71e-16$), Porcentaje Caldenal ($\text{Pr}(> F) = 9.81e-11$), Porcentaje Indice ($\text{Pr}(> F) = 0.000984$). La variable Carga ($\text{Pr}(> F) = 0.609$) no resulta significativa en la discriminación de los grupos [16] [17].

Los dos grupos descartados para los análisis económicos presentaron como características diferencial con respecto de los tres que finalmente se seleccionaron un alto valor en la variable Carga. El análisis particular de las características de estos casos indica que corresponden a empresas agropecuarias que realizan encierro a corral

de los bovinos, incrementando notoriamente la carga animal expresada por unidad de superficie.

Los resultados obtenidos para cada grupo se analizaron con expertos del negocio [8] [16] los cuales validaron que las características de los mismos correspondieran a prácticas factibles en las microrregiones y que las relaciones de superficie y actividades son coherentes.

Para resumir las características de cada sistema productivo se calculó el promedio de cada variable disponible sobre superficie y cantidad de cabezas, aunque las mismas no hayan sido utilizadas en la generación del modelo.

La tabla 6 resume los componentes de cada uno de los grupos seleccionados (cantidad de productores que involucra, superficie y su distribución), mientras que la tabla 7 muestra las principales variables que describen el uso de la tierra, la carga ganadera, la nomenclatura final asignada a los sistemas resultantes y el MB como indicador económico general del sistema.

Tabla 6. Grupos resultantes del clustering.

Grupos	Cantidad de Productores	%	Superficie total	%
Grupo 1*	340	28,8	186.704	23,9
Grupo 2*	83	7,0	137.661	17,6
Grupo 3*	685	58,0	422.162	54,0
Grupo 4	72	6,1	34.656	4,4
Grupo 5	2	0,2	535	0,1
Total	1.108	100	746.527	100

* Grupos seleccionados

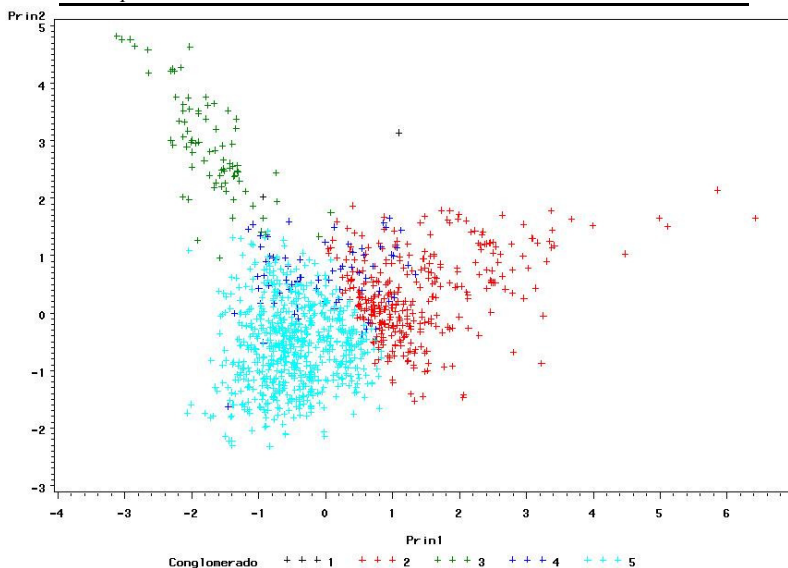


Figura 2. Representación de los clusters por medio de componentes principales. 5 Grupos.

Las características principales de los grupos finales se resumen a continuación:

Grupo 1: la superficie agrícola total superior al 50% permite clasificarlo como sistema “mixto”. La actividad ganadería bovina es de ciclo completo, con preponderancia a la invernada de compra, que se ajusta a los recursos forrajeros disponibles. La carga por superficie ganadera resulta de 0,75 EV/ha. Para este grupo se propone la denominación de sistema “MIXTO CON CICLO COMPLETO CON PREDOMINANCIA INVERNADA”. El mayor resultado económico global está influenciado por la alta rentabilidad actual de la agricultura.

Grupo 2: caracterizado por una importante proporción de pastizal natural, que es consistente con una elevada proporción de vientres y que reduce la carga animal general (0,69 EV/ha). Incluye cultivos forrajeros anuales y perennes que permiten invernar la propia producción. Una escasa superficie agrícola (12%) resulta en una baja superposición en el uso del suelo (5%). Este grupo recibe la denominación “CICLO COMPLETO CON PREDOMINANCIA CRÍA”.

Grupo 3: corresponde al grupo más representativo de la zona (58% de los productores y 54% de la superficie). La oferta forrajera genera la mayor carga animal y permite una cría con invernada de la propia producción y animales de compra, en igual proporción. Se lo ha definido como “CICLO COMPLETO”.

Tabla 5. Indicadores técnico-económicos de los sistemas preponderantes.

Cluster Variables	G 1: Mixto ciclo completo c/ predominancia invernada	G 2: Ciclo completo c/ predominancia cría	G 3:Ciclo completo
Superficie promedio (ha)	550	1660	620
Agricultura (ha)	382	212	188
Trigo	139	40	35
Girasol	129	118	109
Maíz	46	54	44
Soja	68	0	0
Verdeos de Invierno (ha)	85	90	100
Avena	50	50	70
Centeno	35	40	30
Verdeos de Verano (ha)	40	146	86
Maíz	25	65	60
Sorgo	15	81	26
Pasturas Perennes (ha)	80	85	106
Alfalfa	80	85	106
Natural (ha)	115	1205	200
Bosque de Caldén	115	1092	155

(“Caldenal”)			
Campo “Pampa”	0	113	45
Sup. utilizable (ha)	702	1732	680
Superposición	28%	5%	10%
Carga (E.V./ha)	0,75	0,69	0,81
% Agricultura	54%	12%	28%
% Forrajera	46%	88%	72%
MB del sistema (\$/ha)¹	971	772	850

La figura 3 presenta las variables utilizadas para el cálculo de los indicadores técnicos-económicos y cuales permiten distinguir entre los grupos. La figura 4 muestra los diagramas de caja de las siete variables utilizadas en el modelo, donde se aprecia la diferencia de las medias de las variables en cada grupo, permitiendo diferenciar cuales separan un grupo de otro.

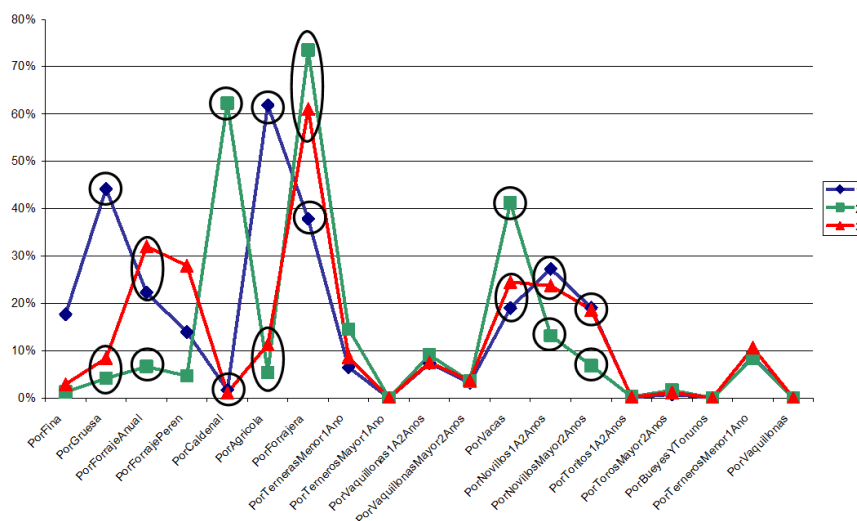
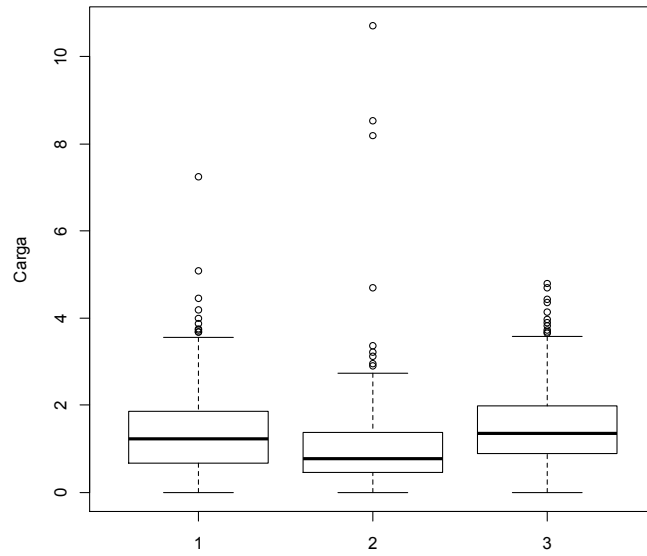


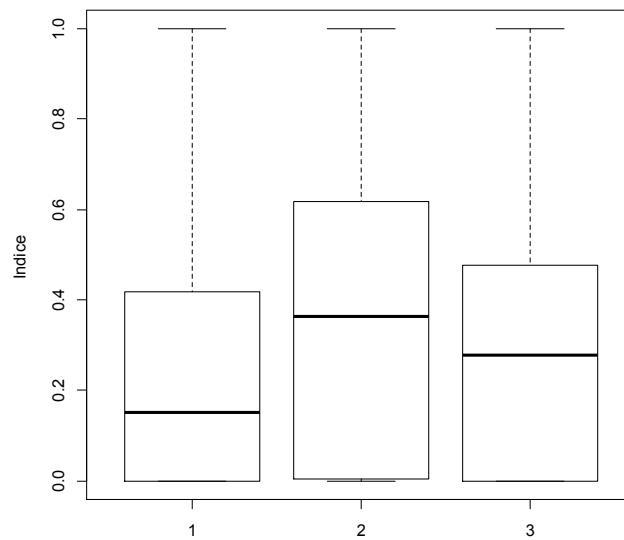
Figura 3. Diferencias entre las variables de los tres grupos finales.

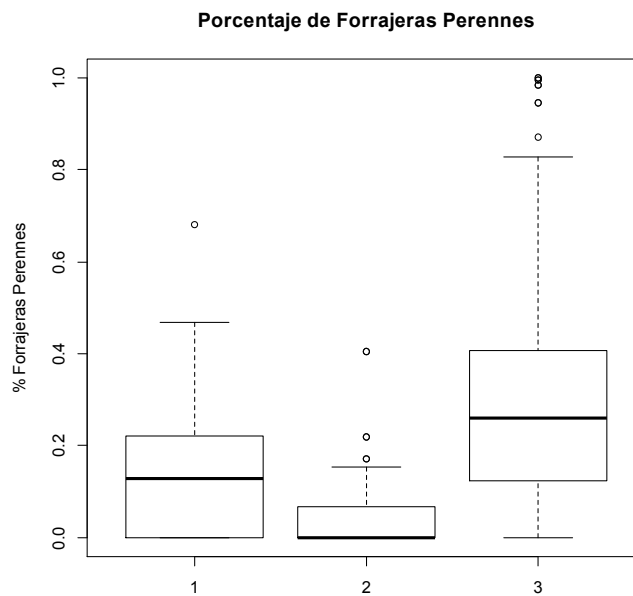
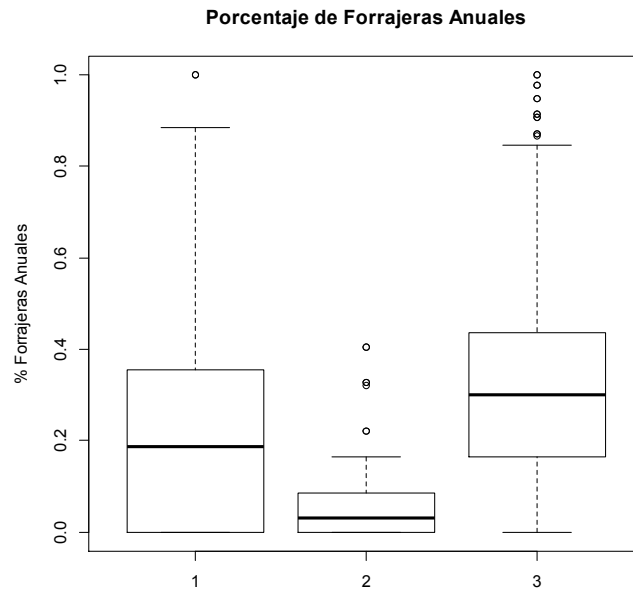
¹ Valores en pesos de septiembre de 2011.

Porcentaje de Carga

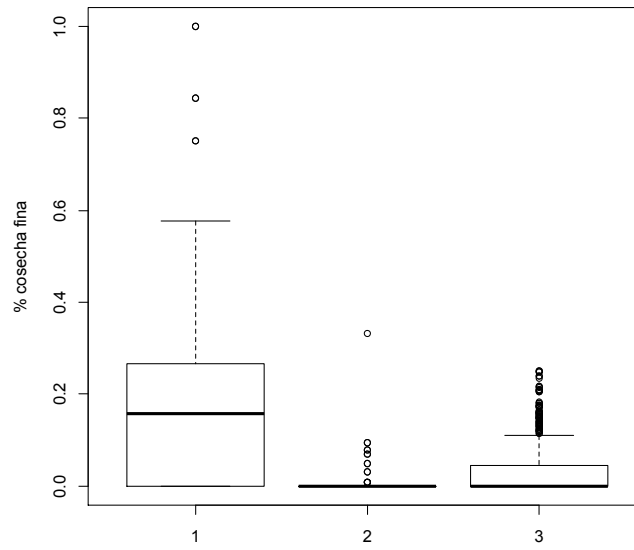


Porcentaje de Indice

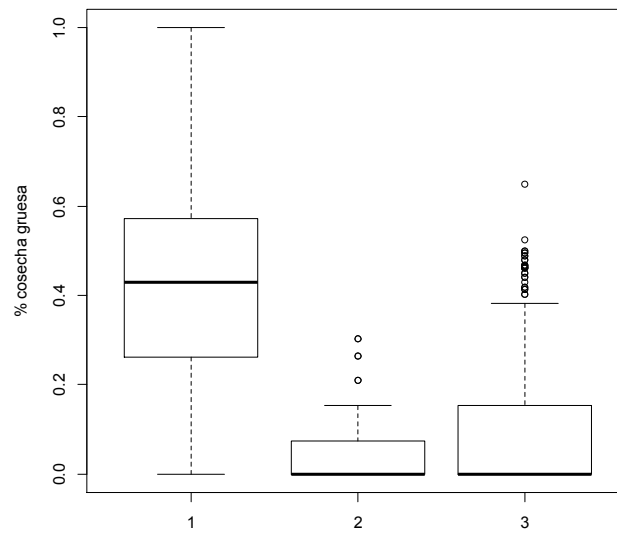




Porcentaje de Cosecha Fina



Porcentaje de Cosecha Gruesa



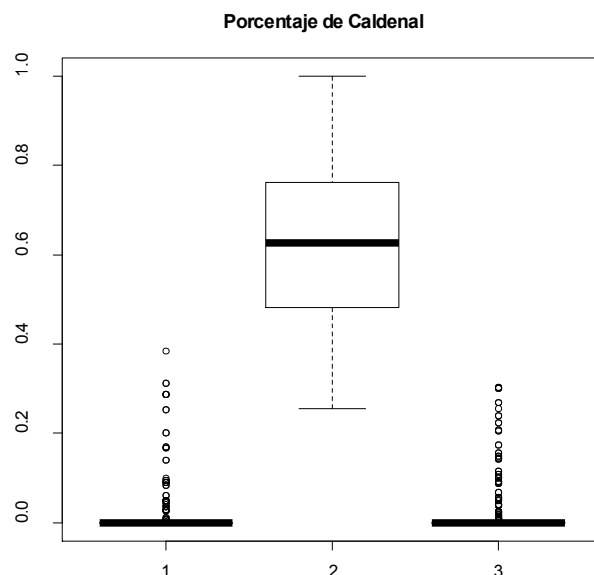


Figura 4. Diagrama de cajas por grupo de las siete variables utilizadas en el armado de los grupos.

4 Conclusiones

La aplicación de técnicas de clustering es una herramienta adecuada para la identificación y caracterización técnica-económica de sistemas productivos preponderantes, para zonas agroecológicas homogéneas.

El modelo generado contiene pocas variables (siete en total) por lo que resulta en un modelo sencillo y se facilita su implementación. Se podrían reducir a seis si se retira la variable carga animal, que no resulta significativa para discriminar grupos.

Cualquier variable técnico productivas a ingresar en el modelo debe ser transformada para expresarla como proporción de la superficie de la explotación agropecuaria para posteriormente ser estandarizada. Solo de esta manera las variables expresan valores que permiten separar grupos.

La disponibilidad y acceso a base de datos de relevamiento anual, permitirá replicar esta metodología a todo el ámbito provincial y analizar su dinámica estructural y económica en forma temporal.

Bibliografía

1. Durán, R.; Scoponi, L. y colaboradores. El gerenciamiento agropecuario en el siglo XXI. Bases para una competitividad sustentable. ISBN 987-1140-31-2. (2005)

2. Iglesias, D.; Iturrioz, G.; Lorda, H.; Fernandez, D.; Leonhardt, D.; Manglus, L.; Rocco, E.: Boletín Económico. 3º Informe Anual. Márgenes brutos de los principales productos agropecuarios de la provincia de La Pampa. INTA Anguil. (Septiembre de 2011).
3. Ghida Daza, C. Indicadores económicos para la gestión de empresas agropecuarias. Bases metodológicas. Informe N° 11. ISSN 1851-6955. INTA.
4. Iglesias, D.; Manazza, F.; Belgrano Rawson, A.; Garay, J.; Bonatti, R.; Lopez, G. Costos de producción y márgenes brutos de los principales productos agropecuarios de la Provincia de San Luis, por regiones. INTA San Luis. (Junio 2012).
5. Caviglia, J.A.; Lorda, H.O.; Lemes, J.D. Caracterización de las unidades de producción agropecuarias en la provincia de LA PAMPA. Boletín de Divulgación Técnica N° 99. Ed. EEA INTA Anguil, 43 pp. Diciembre de 2010. ISSN 03252167
6. Smith, R.; Moreira, V.; Latrille, L. Caracterización de sistemas productivos lecheros en la X región de Chile mediante análisis multivariante. Agric. Téc. v.62 n.3 Chillán jul. 2002. ISSN 0365-2807.
7. Brescia, V.; Rabaglio, M. Atlas informático de los sistemas de producción agropecuarios de Argentina. INTA Instituto de Economías y Sociología. Septiembre 2010.
8. Calvi, M.; Rodríguez, M.; Pueyo, J. M.; Sampietro, D.; Dupleich, J.; Pizzio, R. Caracterización productiva y económica del sistema de cría vacuna predominante en la región centro sur de Corrientes y centro norte de Entre Ríos. INTA. N° 477. Septiembre de 2011. ISSN N° 0327-3059.
9. Demey, J. R.; Adams, M.; Freitas, H. Uso del método de análisis de componentes principales para la caracterización de fincas agropecuarias. Agronomía Tropical. 44(3): 475-497.1994.
10. Lorda, H.; Roberto, Z.; Bellini Saibene, Y.; Sipowicz, A.; Belmonte, M. Descripción de las zonas y subzonas agroecológicas RIAP. Área de influencia de la EEA Anguil. Boletín de Divulgación Técnica N° 96. Octubre de 2008. ISSN 0325-2167.
11. Leyes 468 y 982 de Unidades Económicas de la provincia de La Pampa. Dirección de Economía Agropecuaria. Ministerio de Asuntos Agrarios. Provincia de La Pampa.
12. Las Micro-regiones Pampeanas. Territorios para un nuevo desarrollo. Ministerio de la producción de la provincia de La Pampa, <http://www.produccion.lapampa.gov.ar/libroplaneamiento/microregiones.htm>
13. D. J. Hand, H. Mannila, and P. Smyth, Principles of data mining. The MIT press, 2001.
14. Witten, I.H.; Frank, E.; Hall M. A. Data Mining. Practical Machine Learning Tools and Techniques. Third Edition. (2011).
15. Berzal, F. Clustering. DECSAI. Universidad de Granada. 2006.
16. Justel, A. Técnicas de análisis multivariante para agrupación. Universidad Autónoma de Madrid. 2011.
17. Perea, J. Análisis multivariante para investigación en sistemas agropecuarios. Universidad de Córdoba. 2010.
18. Escobar, G.; Berdegue, J. Tipificación de sistemas de producción agrícola. Septiembre 1990. ISBN. 956.7110-01-07.