METODOLOGIA ESTADISTICA PARA LA CARACTERIZACION DE FINCAS DE CUYES (*CAVIA PORCELLUS*)[1]

M. Zaldívar A.[*], *C. Menacho Ch.*[**]

ABSTRACT

The main objective of this study was to evaluate a statistical methodology for the characterization of guinea pig farms (*Cavia porcellus*). The methodology is based on the theory of principal components, which analyzes the structural dependency of sets of multivariate data. In order to properly evaluate the procedure, information from surveys conducted in Cajamarca, Peru, in 1987 and 1988, was used. The analyses were carried out with the SPSS/PC package. The results of the study proved that the original nine relevant variables could be substituted by four principal components which explained 65% of the total variation as compared to the original variables. The first principal component expresses the livestock productivity and farm income derived from the monthly sale of guinea pigs. The second component reflects the behaviour of the head of the family according to age and schooling. The third component examines the family farm labor that could be supported by the monthly consumption of guinea pigs. The fourth explains that, as family labor becomes scarce, on-farm consumption of guinea pigs also increases. Besides allowing easy, precise characterization of production systems, the principal component analysis identifies the more important variables and their degree of association; as an example of its usefulness, it was found that guinea pig production is, for the most part, geared for self-consumption but this increases if family labor availability decreases, which negatively affects family income.

(**Palabras claves:** Análisis de componentes principales, diagnóstico de fincas, enfoque de sistemas, cuyes).

COMPENDIO

El objetivo principal del presente estudio fue evaluar una metodología estadística para la caracterización de fincas de cuyes (*Cavia porcellus*). Esta se sustenta en la teoría de componentes principales, que analiza la dependencia estructural de conjuntos de datos multivariados. Para evaluarla se utilizó la información de las encuestas de caracterización de los sistemas de producción de cuyes, realizadas en la Provincia de Cajamarca, Perú, entre 1987 y 1988. Para el análisis se usó el paquete *Statistical Package for Social Scientists* (SPSS)/PC. Los resultados de la aplicación muestran que las nueve variables relevantes, consideradas inicialmente en el presente caso, pueden ser reemplazadas por cuatro componentes principales que explican el 65% de la variación total respecto de las originales. El primer componente principal expresa la productividad pecuaria e ingreso de la finca por la venta mensual de cuyes. El segundo refleja el comportamiento del jefe de familia considerando su edad y grado de instrucción. El tercero sintetiza la mano de obra familiar -disponible y no disponible- que puede tener la finca con el consumo mensual de cuyes. El cuarto componente explica el consumo mensual de cuyes y su efecto negativo -disminución- al incrementarse la mano de obra familiar no disponible. Además de permitir la caracterización fácil y precisa de los sistemas de producción, el análisis de componentes principales identifica las variables de mayor importancia y sus asociaciones entre sí; como ejemplo de esto último, se encontró que la producción de cuyes se dirige principalmente al autoconsumo, pero que éste se incrementa con el aumento de la mano de obra familiar no disponible, lo que atenta contra el ingreso de la familia.

INTRODUCCION

Todos los trabajos sobre tipificación y caracterización de fincas coinciden en considerarlas como un sistema en el cual interactúan diferentes tipos de recursos, procesos y fines productivos con un orden jerárquico para generar los subsistemas. Es necesario precisar conceptos de sistemas, subsistemas y sus componentes a fin de tipificar y caracterizar adecuadamente un conjunto de productores.

Hotelling (1) fue el primero en formular el análisis de componentes principales, basándose en el trabajo publicado en 1901 por Karl Pearson sobre el ajuste de un multiespacio a una línea o a un plano. El enfoque de Pearson se centra en el análisis de componentes que sintetiza la mayor variabilidad del sistema de puntos; ello explica el calificativo de "principal".

[1] Recibido para publicación el 18 de marzo de 1991.

[*] Instituto Nacional de Investigación Agraria y Agroindustrial, La Molina, Lima, Perú.

[**] Universidad Nacional Agraria La Molina, Lima, Perú.

El análisis de componentes principales es un método que examina la dependencia estructural de datos multivariados obtenidos de una población, cuya distribución de probabilidades no es preciso conocer. Sin embargo, puede suponerse que la población muestreada tiene distribución multinormal, con lo que se podrían realizar las respectivas pruebas de hipótesis para extraer inferencias de la población en estudio.

Con el análisis de componentes principales se pretende generar nuevas variables que expresen la mayor parte de la información contenida en el conjunto original, reducir el número de variables para una mejor interpretación de los datos, y eliminar aquellas que aportan escasa información (2, 3).

El objetivo de la metodología propuesta fue evaluar la técnica multivariada del análisis de componentes principales en la caracterización de fincas de producción de cuyes.

METODOLOGIA

Teniendo en cuenta que se pueden definir los componentes del sistema de finca como un conjunto de variables que se interrelacionan, es posible inferir mediante el análisis de los componentes principales la dependencia estructural de dichas variables y, por consiguiente, aquellos que determinan el sistema productivo de las fincas.

Para evaluar y probar la metodología se usó la información de 83 encuestas de caracterización de producción de cuyes, realizada en la Provincia de Cajamarca, Perú, entre 1987 y 1988, procesándola mediante el paquete de análisis estadístico para microcomputadoras SPSS/PC versión 1.1.

Un conjunto de datos constituye una muestra aleatoria multivariada de tamaño **n**, si se han extraído **n** individuos de una población con vector de medias y matriz de variancias-covariancias δ, y en ella se han medido u observado **p** características (variables). Sean X_{ij} la observación correspondiente a la j-ésima variable en el i-ésimo individuo, X_i el vector fila que contiene las observaciones de todas las variables en el i-ésimo individuo y X_j el vector columna que contiene todas las observaciones de la j-ésima variable. Se define la matriz de datos multivariados X, de dimensión (**nxp**), como el arreglo:

$$X = \begin{bmatrix} X_{11} & X_{12} & X_{13} & \cdots & X_{1p} \\ X_{21} & X_{22} & X_{23} & \cdots & X_{2p} \\ X_{31} & X_{32} & X_{33} & \cdots & X_{3p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{n1} & X_{n2} & X_{n3} & \cdots & X_{np} \end{bmatrix}$$

Dada la matriz de datos multivariados X, se define la media muestral de la j-ésima variable por el vector fila (1xp):

$$\bar{x}_j = \frac{1}{n} \overset{n}{\underset{i=1}{\delta}} X_{ij}, \quad \text{para } j = 1,2,3,\ldots,p$$

y la variancia muestral de la j-ésima variable S_{jj} y covariancia de la j-ésima y k-ésima variables S_{jk}, por:

$$S_{jj} = \frac{1}{n-1} \overset{n}{\underset{i=1}{\delta}} (X_{ij} - X_j)^{\cdot 2}$$

$$S_{jk} = \frac{1}{n-1} \overset{n}{\underset{i=1}{\delta}} (X_{ij} - X_j)^{\cdot} (X_{ik} - X_k)^{\cdot}$$

para j,k = 1,2,3,...,p

Los valores S_{jj} y S_{jk} serán los correspondientes elementos de la matriz de variancias-covariancias muestral S de dimensión (pxp), esto es:

$$S = \begin{bmatrix} S_{11} & S_{12} & S_{13} & \cdots & S_{1p} \\ S_{21} & S_{22} & S_{23} & \cdots & S_{2p} \\ S_{31} & S_{32} & S_{33} & \cdots & S_{3p} \\ \vdots & \vdots & \vdots & \vdots & \\ S_{p1} & S_{p2} & S_{p3} & \cdots & S_{pp} \end{bmatrix}$$

A partir de los elementos de la matriz S, es posible calcular los coeficientes de correlación simple R_{jk} de la j-ésima variable y k-ésima variable, cuyos valores serán los elementos de la matriz de correlaciones R de dimensión (pxp), esto es:

$$r_{jk} = \frac{S_{jk}}{1\,S_{jj} \quad 1\,S_{ij}}$$

$$R = \begin{array}{ccccc} 1 & & & & \\ r_{21} & 1 & & & \\ X_{31} & r_{32} & 1 & & \\ \vdots & \vdots & \vdots & & \\ r_{p1} & r_{p2} & r_{p3} & \cdots & r_{pp} \end{array}$$

La matriz S expresa la dispersión de los datos alrededor de la media. A veces, es necesario encontrar un número o escalar que sintetice completamente la variabilidad de los datos multivariados a partir de la información contenida en la misma matriz S. A partir de la matriz de variancias-covariancias S, se calcula y define la variancia generalizada V al determinante de dicha matriz, es decir:

$$V = |S|$$

calculándose la variancia total (VT) a la traza de la matriz S, esto es:

$$VT(X) = Tr(S) = \delta_{j=1}^{p} S_{jj}$$

Tanto la variancia generalizada como la total serán mayores cuanto mayor sea la dispersión de los datos alrededor de la media. Sin embargo, cada medida refleja aspectos diferentes de la variabilidad de los datos. La primera desempeña un papel importante en la generación de los estimadores máximos verosímiles, mientras que la segunda se utiliza para el análisis de componentes principales.

Es posible encontrar un escalar $\mathbf{L}_j$ y un vector $\mathbf{W}_j$ de dimensión (px1) no nulo a partir de la matriz cuadrada B de dimensión (pxp), tal que:

$$BW_j = L_jW_j$$

lo que implica que:

$$(B - L_jI)W_j = 0$$

Según el álgebra lineal, la solución de la expresión anterior es cuando: $|B - L_jI| = 0$, conduciendo a un polinomio de grado **p** cuyas soluciones son los autovalores $\mathbf{L}_j$ y los autovectores $\mathbf{W}_j$ de la matriz B.

Este método multivariado se basa en encontrar una transformación lineal de un conjunto de variables originales a un nuevo conjunto de variables ortogonales (no correlacionadas), denominadas componentes principales; tal que sus variancias se maximicen en orden decreciente.

Al estudiar un conjunto de **n** individuos mediante **p** variables originales $\mathbf{X}_j$, j=1,2,3,...,p; es posible encontrar nuevas variables $\mathbf{Y}_k$, k=1,2,3,...,p; que sean combinaciones lineales de la $\mathbf{X}_j$ originales, tal que este nuevo conjunto (componentes principales) cumpla con las siguientes condiciones:

1) que los componentes principales no estén correlacionados

$$Cov(X_j, X_k) = 0 \quad , \quad \text{para } j \pm k;$$

2) que cada componente principal sintetice la máxima variabilidad en orden decreciente

$$Var(Y_1) \geq Var(Y_2) \geq Var(Y_3) \geq ... \geq Var(Y_p) \geq 0$$

Esto implica entonces encontrar (pxp) constantes tales que:

$$Y_k = \delta_{j=1}^{p} W_{jk}X_j \text{ , para } k = 1,2,3,...,p$$

donde los $\mathbf{W}_j$ son esas constantes y sujetas a la condición que sean vectores ortogonales, esto es:

$$W_j'W_k = \begin{bmatrix} 1 \ , & \text{si } j = k \\ 0 \ , & \text{si } j \neq k \end{bmatrix}$$

Usando los parámetros poblacionales de las variables $\mathbf{X}_j$, es posible encontrar los parámetros poblacionales de los componentes principales $\mathbf{Y}_j$, es decir:

$$E(Y_j) = \mu X$$
$$Var(J_j) = W_j'\delta W_j$$

La forma general de la función que se ha de maximizar, aplicando los estimadores de Lagrange para estimar el j-ésimo componente principal (Y_j), será:

$$H(W_j, L_j) = W_j'\delta W_j - L_j(W_j'W_j - 1)$$

donde los $\mathbf{L}_j$ son los multiplicadores de Lagrange. Entonces:

$$\frac{U\, H(W_j,L_j)}{U\, W_j} = 2\delta W_j - 2\, L_jW_j = 0$$

$$\delta W_j = L_jW_j$$

lo que implica que:

$$(\delta - L_jI)W_j = 0$$

La estimación del j-ésimo componente principal está dada por el cálculo de los autovalores $\mathbf{L}_j$ y sus correspondientes autovectores $\mathbf{W}_j$, respecto de la matriz de variancia-covariancia δ; entonces, se define el j-ésimo componente principal como:

$$Y_j = W_j'X$$
$$Var(Y_j) = W_j'SW_j = L_j$$

donde:

L_j = Son los autovalores de la matriz de variancia-covariancia δ. La estimación de δ es S.

W_j = Son vectores columnas correspondientes a los autovectores de δ.

La matriz de componentes principales **Y**, de orden (nxp), se define como la transformación lineal respecto de la matriz de datos **X** de orden (nxp) que es diagonalizada por la matriz de autovectores **W** de orden (pxp):

$$Y = X\, W$$

Para la interpretación de los componentes se consideraron tres etapas:

a .Selección del número de componentes principales

Como cada componente principal Y_j es una transformación lineal ortogonal de las variables X_j, entonces la variación total de las X_j es igual al de las Y_j:

$$\sum_{j=1}^{p} S_{jj} = \sum_{j=1}^{p} L_j$$

Cada componente principal explica una proporción de la variación total. Esta proporción, llamada proporción de variación, es explicada por el j-ésimo componente [PVE(Y_j)] que se calcula por:

$$PVE(Y_j) = \frac{L_j}{\sum_{j=1}^{p} L_j}$$

y la proporción de variación explicada acumulada por los **k** primeros componentes principales (PVEA), se calcula por la suma de los PVE(Y_j) hasta **k**:

$$PVEA = \sum_{j=1}^{k} PVE\ (Y_j)$$

La PVEA multiplicada por 100 indica el porcentaje de proporción de variación explicado por los **k** primeros componentes. Esto fue usado para decidir el número de componentes principales que se seleccionaron. El criterio de Kaiser se utilizó para seleccionar aquellos componentes cuyos autovalores fueron mayores o iguales al promedio.

b. Prueba de hipótesis

Las hipótesis planteadas correspondientes serán:

$$Ho\ :\ \frac{L_1 + L_2 + \ldots + L_k}{L_1 + L_2 + \ldots + L_p} = a$$

$$Ha\ :\ \frac{L_1 + L_2 + \ldots + L_k}{L_1 + L_2 + \ldots + L_p} < a$$

Siendo la prueba estadística:

$$Zc = \frac{(a - \hat{a})}{T} \sim N(0, 1)$$

donde: $E(a) = a$

$$Var(a) = T^2 = \frac{2\ Tr\ (\Sigma^2)}{(n-1)\ (Tr\Sigma)^2}\ (a^2 - 2ea + e)$$

y las estimaciones muestrales:

$$Tr\Sigma = TrS = \sum_{j=1}^{p} l_j \quad y \quad Tr\Sigma^2 = TrS^2 = \sum_{j=1}^{p} l_j^2$$

$$\hat{e} = \sum_{j=1}^{k} l_j^2 / \sum_{j=1}^{p} l_j^2 \quad y \quad \hat{a} = \sum_{j=1}^{k} l_j / \sum_{j=1}^{p} l_j$$

$$t^2 = \frac{2TrS^2}{(n-1)\ (TrS)^2}\ (\hat{a} - 2\hat{a}\hat{e} + \hat{e})$$

c. Correlación entre las variables originales y los componentes principales

La correlación $r(X_j, Y_j)$ representa el grado de asociación existente entre las variables originales X_j y los componentes Y_j. Así, se tiene:

$$r(x_j, Y_j) = \frac{W_{ij}\ l_j}{S_{jj}}$$

se cumple que:

$$\sum_{Y_j=1}^{p} r^2\ (X_j, Y_j) = 1$$

RESULTADOS Y DISCUSION

Se seleccionaron las variables siguientes, cuyos promedios, desviaciones estándar y coeficientes de variabilidad se presentan en el Cuadro 1.

Las variancias y covariancias de las variables originales (Cuadro 2) presentan un rango amplio de valores debido a la influencia de las magnitudes absolutas (años, número de animales, otros). En las covariancias el signo indica la existencia de una dependencia directa o indirecta.

Ya que las variables de la matriz "S" están en diferentes unidades, para el análisis se utilizó la matriz de correlaciones descrita en el Cuadro 3.

En el Cuadro 4 se presentan los autovalores asociados a la matriz "R" de correlaciones. Se muestran las variancias de los componentes, dados por sus correspondientes autovalores. El primer componente explica el 24.2% de la variación total. Los cuatro primeros componentes principales, que tienen autovalores mayores que el promedio (superiores que 1), dan un porcentaje de PVEA de 64.5% y fueron seleccionados según el criterio de Kaiser.

Cuadro 1. **Variables consideradas en el estudio.**

No.	Nombre	Código	Promedio	Des.Est.	Coef.Var.
1	Edad jefe de familia	VAR1	35.4	25.38	59.92
2	Años de estudio del jefe de familia	VAR2	2.8	3.19	112.68
3	Mano de obra familiar disponible	VAR3	1.8	1.12	60.22
4	Promedio de otros animales	VAR4	2.2	2.14	98.12
5	Número de cuyes actual	VAR5	22.6	25.66	113.33
6	Promedio de cuyes por calidad	VAR6	8.9	10.66	83.13
7	Venta mensual de cuyes por tamaño	VAR7	12.8	44.04	343.58
8	Consumo mensual por tamaño	VAR8	30.2	70.80	234.11
9	Mano de obra familiar disponible	VAR9	0.6	0.87	144.35

VAR1: Edad del esposo o del responsable de la finca.

VAR2: Años de estudio del esposo, considerando las categorías PI=2.5, PC=5, SI=7.5, SC=10 y S=15 años de estudio.

VAR3: Número de personas de la finca cuya ocupación sea en su casa, agricultura o ganadería.

VAR4: Número promedio de otros animales criados en la finca (vacunos, ovinos, porcinos, equinos, conejos y aves).

VAR5: Número total de cuyes que tiene la finca.

VAR6: Promedio ponderado de cuyes reproductores, de recría y lactantes; cuya ponderación está dada por la condición en que se encuentran (bueno = 6, regular = 4 y malo = 2).

VAR7: Número de cuyes vendidos mensualmente (ingreso) ponderado por su tamaño (grande = 6, mediano = 4, chico = 2 e indistinto = 4).

VAR8: Número de cuyes consumidos mensualmente (egreso), ponderado por su tamaño (grande = 6, mediano = 4, chico = 2 e indistinto = 4).

VAR9: Número de personas de la finca cuya ocupación (empleado, comerciante, artesano, profesional, etc.) es otra que en su casa, agricultura o ganadería

Cuadro 2. **Matriz "S" de variancias-covariancias.**

	VAR1	VAR2	VAR3	VAR4	VAR5	VAR6	VAR7	VAR8	VAR9
1	643.882								
2	24.031	10.179							
3	6.124	-0.134	1.247						
4	3.646	1.754	0.528	4.591					
5	26.268	-0.671	2.081	25.747	658.282				
6	27.743	0.027	1.397	5.243	171.130	113.549			
7	35.762	-13.506	2.242	9.568	217.446	138.119	193.906		
8	-90.764	-0.386	-14.306	-4.212	-149.863	-76.366	75.215	5 012.039	
9	-1.325	0.347	-0.156	-0.199	-2.694	-0.934	-5.036	-0.391	0.775

Cuadro 3. **Matriz "R" de correlaciones.**

	VAR1	VAR2	VAR3	VAR4	VAR5	VAR6	VAR7	VAR8	VAR9
1	1								
2	0.2968*	1							
3	0.2161*	-0.0377	1						
4	0.0671	0.2567*	0.2210*	1					
5	0.0403	-0.0082	0.0726	0.4684*	1				
6	0.1026	0.0008	0.1174	0.2297*	0.6259*	1			
7	0.0320	-0.0961	0.0456	0.1014	0.1924	0.2943*	1		
8	-0.0505	-0.0017	-0.1810	-0.0278	-0.0825	-0.1012	0.0241	1	
9	-0.0601	0.1252	-0.1605	-0.1071	-0.1209	-0.1009	-0.1316	-0.0064	1

83 observaciones
* Significativo ($P \leq 0.05$)

Cuadro 4. Autovalores de la matriz de correlaciones.

Componente	Autovalor	Proporción de variación explicada Absoluta (%)	Acumulada (%)
1	2.18073	24.2	24.2
2	1.39694	15.5	39.8
3	1.19089	13.2	53.0
4	1.03438	11.5	64.5
5	0.92923	10.3	74.8
6	0.76367	8.5	83.3
7	0.73192	8.1	91.4
8	0.49295	5.5	96.9
9	0.27928	3.1	100.0

Cuadro 5. Correlación (r^2) entre las variables originales con los componentes principales.

Variable	Componentes principales 1	2	3	4	Σr^2
VAR1	0.0709	0.3911	0.0475	0.0999	0.6093
VAR2	0.0156	0.5946	0.1372	0.0114	0.7588
VAR3	0.1434	0.0557	0.4466	0.0004	0.6460
VAR4	0.4202	0.0605	0.0448	0.0000	0.5255
VAR5	0.6474	0.0337	0.0779	0.0290	0.7880
VAR6	0.5782	0.0469	0.0289	0.0164	0.6704
VAR7	0.1746	0.1438	0.0000	0.1046	0.4230
VAR8	0.0385	0.0199	0.1856	0.5329	0.7769
VAR9	0.0921	0.0508	0.2224	0.2397	0.6050

Para la decisión definitiva del número de componentes seleccionados, fue necesario calcular los autovalores y las correlaciones de las variables originales con los componentes principales. Los autovalores dieron la ponderación que generó los componentes principales, indicando el grado de importancia de la correspondiente variable; mientras que las correlaciones explicaron la proporción de los componentes respecto a cada una de las variables.

Los coeficientes de correlación entre las variables originales y los componentes principales se calcularon al dividir cada elemento de la matriz de autovectores por su respectivo autovalor. La sumatoria de las correlaciones con los cuatro componentes principales, elevados al cuadrado, explican la variación para cada variable original debido a los componentes seleccionados (Cuadro 5).

Según el Cuadro 6, el primer autovector dio una ponderación - promedio de todas las variables para el primer componente principal, con valores mayores para las variables 4, 5 y 6 (0.4389, 0.5449 y 0.5149) y menores para las variables 3, 7 y 1; la contribución de la variable 2 es casi nula. El valor -0.1328 indica que disminuirá el valor del primer componente si aumenta el valor de la variable 8 (consumo mensual). El valor -0.2055 manifiesta una disminución de la productividad pecuaria de la finca si disminuye la mano de obra no familiar.

El primer componente (véase el Cuadro 5) refleja la composición del hato y el plantel de cuyes. En el Cuadro 5 se puede apreciar que las mayores proporciones de explicación (correlación) del primer componente corresponden a las variables 4, 5 y 6 (42.02%, 64.74% y 57.82%), respectivamente y, en menor importancia, a las variables 2, 3 y 7. Las fincas con valores elevados para el primer componente principal están asociadas a la mejor composición del hato y plantel de cuyes, regular ingreso por venta de cuyes, bajo consumo mensual de cuyes e insignificante disminución en la productividad pecuaria si aumenta la mano de obra no disponible.

El segundo autovector (Cuadro 6) da las mayores ponderaciones positivas para las variables 1 y 2 (0.5291 y 0.6524), y las menores para las variables 3 y 4. Coeficientes negativos, particularmente para la variable 7 (-0.3218) pero también para las variables 5, 6 y 8. Estos valores negativos indican que el segundo componente disminuirá, no significativamente, si la variable 7 se incrementa y, en forma aún menor, si aumentan los valores de las variables 5, 6 u 8. Del Cuadro 5 se deriva que el segundo componente principal explica el comportamiento del jefe de familia en función de su edad y nivel de escolaridad. Los mayores valores para el segundo componente se dan en las variables 1 y 2 (39.11% y 59.46%); también se observa que las ventas de cuyes son perjudiciales, pero no significativamente, para el jefe de familia.

Cuadro 6. Matriz de autovectores de la matriz R.

Variable	1	2	3	4
VAR1	0.1803	0.5291	-0.1997	0.3108
VAR2	0.0845	0.6524	0.3394	0.1052
VAR3	0.2564	0.1997	-0.6124	-0.0189
VAR4	0.4389	0.2081	0.1940	0.0055
VAR5	0.5449	-0.1553	0.2558	-0.1675
VAR6	0.5149	-0.1832	0.1559	-0.1259
VAR7	0.2829	-0.3208	0.0026	0.3180
VAR8	-0.1328	-0.1193	0.3948	0.7178
VAR9	-0.2055	0.1907	0.4321	-0.4814

Para el tercer autovector (Cuadro 6) las variables 8, 9 y 3 presentan los mayores valores, siendo la última variable la de mayor valor absoluto con un coeficiente negativo (-0.6124). Esto indica que el tercer componente disminuirá significativamente si se incrementa el valor de la variable 3. Las fincas con valores elevados para el tercer componente principal (Cuadro 5), estarán asociadas con menor mano de obra familiar disponible y consumo regularmente alto, posiblemente debido a la mano de obra familiar no disponible.

Los valores del cuarto autovector (Cuadro 6) presentan coeficientes altos para las variables 8, 7, 1 y 9 (0.7178, 0.3180, 0.3108 y -0.4814), respectivamente. La variable 9 tiene un coeficiente negativo, indicando que el cuarto componente disminuirá al aumentar el valor de esta variable —mano de obra familiar no disponible. Al observar la información del Cuadro 5, el cuarto componente principal explica, en gran medida, el consumo mensual de cuyes y su disminución al aumentar la mano de obra familiar no disponible. Las fincas con valores altos están asociadas a un mayor consumo mensual de cuyes por parte de la mano de obra familiar no disponible.

Las fincas con valores altos del primer componente principal están asociadas con la mejor producción pecuaria —composición del hato y plantel de cuyes—, regular ingreso por la venta mensual de productos, bajo consumo de cuyes y disminución de la productividad (no significativa), si se incrementa la mano de obra familiar no disponible.

El segundo componente principal explica con mayor significación las variables que definen la edad y el grado de instrucción del jefe de familia, y un efecto negativo si se incrementa la venta mensual de cuyes. Por este componente fue también posible formar grupos de fincas —productores—, considerando los valores de este segundo componente.

El destino de la producción del conjunto de productores fue explicado por el tercer componente principal, agrupando las fincas que tuvieron el mayor valor por este concepto en función de la mano de obra familiar disponible.

El cuarto componente principal explica el consumo mensual de cuyes en función de la mano de obra familiar no disponible, siendo su comportamiento similar al tercer componente.

CONCLUSIONES

Las principales conclusiones del presente trabajo son:

1. El análisis de los componentes principales permite una caracterización fácil y precisa del sistema de finca. Analiza la dependencia estructural del conjunto de variables relevantes, explicando el sistema, subsistema y componentes a través de un número menor de variables.

2. Los productores de cuyes de la provincia de Cajamarca (Perú) pueden ser divididos en tres grupos de acuerdo con su productividad pecuaria e ingresos por la venta mensual de cuyes.

3. El destino de la producción de cuyes es principalmente para el autoconsumo. Se observa un mayor consumo por parte de la mano de obra familiar no disponible que por la disponible —casa, agricultura y ganadería.

LITERATURA CITADA

1. HOTELLING, H.O. 1933. Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology 24:417-441, 498-520.

2. MORRISON, D.F. 1967. Multivariate Statistical Methods. New York, McGraw-Hill. 338 p.

3. PLA, L. 1986. Análisis multivariado: Método de componentes principales. Washington.D.C., OEA, Programa Regional de Desarrollo Científico y Tecnológico. p. 40-49.