

Caracterización de la producción lechera de un distrito

Fuente: Análisis Multivariado. Método de Componentes Principales.
Laura E. Pla, OEA, 1986

- La tabla sintetiza las variables analizadas con objeto de conocer la situación del sector lechero venezolano.

Nombre	Código	Promedio	Desv. Estándar	Coef. De variación (%)
Superficie total de la finca (ha)	SUP	383,51	328,25	86
Nro. total de vacas	VACA	100,13	73,23	73
Índice sanitario	SANI	18,17	4,00	21
Índice de instalaciones	INST	71,16	16,43	23
Índice de maquinaria	MAQ	12,38	21,57	174
Promedio de leche/vaca (l)	PROM	2,20	1,50	68

- Los altos valores de los coeficientes de variación (CV=desv. estándar/media) indican una gran heterogeneidad entre los productores del distrito.
- ¿Cuáles variables influyen más sobre la productividad?
- ¿Qué fincas son las más afectadas?

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

Caracterización de la producción lechera de un distrito

- Se calculó la matriz de covarianza utilizando los datos originales, es decir sin estandarizar los valores.

	SUP	VACA	SANI	INST	MAQ	PROM
	107746,3	11875,63	268,34	389,19	1704,49	-25,3
	11875,63	5362,88	24,44	156,36	293,25	-31,44
S=	268,34	24,44	15,97	10,64	14,02	0,76
	389,19	156,36	10,64	269,93	34,59	3,64
	1704,49	293,25	14,02	34,59	465,3	2,68
	-25,3	-31,44	0,76	3,64	2,68	2,25

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

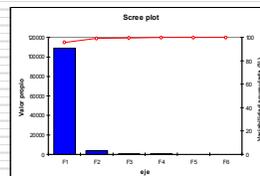
Caracterización de la producción lechera de un distrito

- A partir de la matriz S se obtienen los valores propios y la proporción de la varianza total explicada por cada uno de los componentes.

Análisis de Componentes Principales:

Valores propios:

	F1	F2	F3	F4	F5	F6
Valor propio	109135,344	4009,666	439,169	261,795	14,743	1,913
Variabilidad (%)	95,848	3,521	0,386	0,230	0,013	0,002
% acumulado	95,848	99,370	99,755	99,985	99,998	100,000



- El primer C.P. sintetiza casi el 96% de la varianza total. Esta situación, que a primera vista puede ser muy ventajosa, puede llevar a interpretaciones equivocadas.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

Caracterización de la producción lechera de un distrito

- Es necesario examinar los vectores propios y los coeficientes de correlación de las nuevas variables con las originales.

Vectores propios:

	F1	F2	F3	F4	F5	F6
SUP	0,993	-0,114	-0,013	0,001	-0,002	0,000
VACA	0,114	0,993	-0,033	-0,025	0,003	0,008
SANI	0,002	-0,001	0,026	0,034	0,998	-0,043
INST	0,004	0,030	0,144	0,988	-0,038	-0,015
MAQ	0,016	0,027	0,988	-0,146	-0,022	-0,007
PROM	0,000	-0,007	0,010	0,015	0,042	0,999

- En este ejemplo, cada C.P. puede asociarse con una variable, con aquella que posee el valor mayor en el vector.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

Caracterización de la producción lechera de un distrito

- Esta misma situación se refleja cuando se examinan los valores de la correlación entre las variables y los C.P.

Correlaciones entre las variables y los factores:

	F1	F2	F3	F4	F5	F6
SUP	1,000	-0,022	-0,001	0,000	0,000	0,000
VACA	0,513	0,858	-0,009	-0,006	0,000	0,000
SANI	0,204	-0,023	0,139	0,136	0,959	-0,015
INST	0,075	0,115	0,184	0,973	-0,009	-0,001
MAQ	0,243	0,081	0,960	-0,110	-0,004	0,000
PROM	-0,058	-0,296	0,146	0,166	0,107	0,921

- Estas variables se expresan en distintas unidades, cuyos rangos de valores difieren en uno o más órdenes de magnitud, lo que puede influir notablemente en la interpretación de los resultados.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

Caracterización de la producción lechera de un distrito

- Uso de la matriz de correlación**
a partir de datos estandarizados, donde cada variable tiene media 0 y varianza 1.

	SUP	VACA	SANI	INST	MAQ	PROM
	1	0,494	0,2046	0,0722	0,2407	-0,0514
	0,494	1	0,0835	0,13	0,1856	-0,2865
	0,2046	0,0835	1	0,162	0,1627	0,1265
R=	0,0722	0,13	0,162	1	0,0976	0,1477
	0,2407	0,1856	0,1627	0,0976	1	0,0829
	-0,0514	-0,2865	0,1265	0,1477	0,0829	1

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

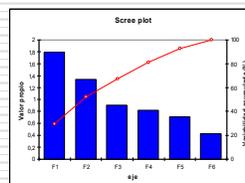
Caracterización de la producción lechera de un distrito

- ❑ Valores propios y proporción de la varianza explicada (cálculos a partir de la matriz R)

Análisis de Componentes Principales:

Valores propios:

	F1	F2	F3	F4	F5	F6
Valor propio	1,794	1,341	0,903	0,821	0,712	0,428
Variabilidad (%)	29,902	22,349	15,053	13,688	11,874	7,134
% acumulado	29,902	52,251	67,304	80,993	92,866	100,000



- ❑ Es necesario considerar hasta 4 C.P. para explicar el 81% de la varianza y hasta 5 C.P. para explicar el 93%.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

Caracterización de la producción lechera de un distrito

Vectores propios:

	F1	F2	F3	F4	F5	F6
SUP	0,587	-0,100	-0,134	0,088	0,534	-0,579
VACA	0,569	-0,342	0,184	-0,062	0,136	0,709
SANI	0,325	0,417	-0,082	0,773	-0,327	0,093
INST	0,242	0,412	0,794	-0,282	-0,156	-0,197
MAQ	0,396	0,229	-0,530	-0,540	-0,467	-0,028
PROM	-0,105	0,691	-0,176	-0,139	0,589	0,338

- F1: es un promedio de todas las variables con ponderación relativa mayor para SUP y para VACA, y algo menor para SANI, INST y MAQ. El coeficiente para PROM es negativo, lo que indica que disminuirá el valor de F1 si aumenta la productividad por vaca. Sin embargo, su contribución es pequeña y las fincas con valores elevados de F1 estarán asociadas a valores elevados de SUP, VACA, SANI, INST y MAQ.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

Caracterización de la producción lechera de un distrito

Vectores propios:

	F1	F2	F3	F4	F5	F6
SUP	0,587	-0,100	-0,134	0,088	0,534	-0,579
VACA	0,569	-0,342	0,184	-0,062	0,136	0,709
SANI	0,325	0,417	-0,082	0,773	-0,327	0,093
INST	0,242	0,412	0,794	-0,282	-0,156	-0,197
MAQ	0,396	0,229	-0,530	-0,540	-0,467	-0,028
PROM	-0,105	0,691	-0,176	-0,139	0,589	0,338

- F2: se observa poca contribución de SUP y coeficientes negativos para SUP y VACA, constituyendo un contraste con las restantes 4 variables. De tal forma que las fincas con los mayores valores de F2 serán aquellas que posean la mayor productividad, las mejores condiciones sanitarias, instalaciones y equipos.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

Caracterización de la producción lechera de un distrito

□ Interpretación de los resultados

Puede graficarse la correlación de cada variable original con los dos primeros C.P.

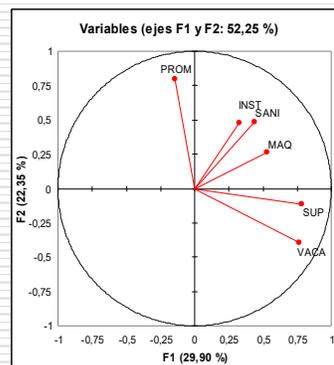
Correlaciones entre las variables y los factores:

	F1	F2	F3	F4	F5	F6
SUP	0,786	-0,116	-0,128	0,080	0,451	-0,379
VACA	0,762	-0,395	0,175	-0,057	0,115	0,464
SANI	0,435	0,483	-0,078	0,701	-0,276	0,061
INST	0,324	0,477	0,754	-0,255	-0,131	-0,129
MAQ	0,531	0,265	-0,503	-0,489	-0,394	-0,019
PROM	-0,141	0,800	-0,168	-0,126	0,497	0,221

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

Caracterización de la producción lechera de un distrito

- El círculo de radio unitario, trazado con centro en el origen del par de coordenadas, permite identificar las variables cuya correlación con los C.P. sea mayor (se acercan al círculo). Aquellas variables que se agrupan cerca del centro serán las que menos se correlacionan con los C.P.



AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

Caracterización de la producción lechera de un distrito

- Además de graficar las variables originales en función de los 2 primeros C.P., es posible graficar los valores de los 2 primeros C.P. para cada observación. Si el gráfico se superpone al anterior se denomina birrepresentación ("biplot" en inglés).

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

Caracterización de la producción lechera de un distrito

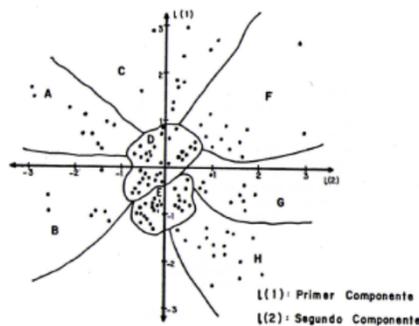


Fig. 5. Representación bivariada de resultados. Cada sector, identificado con una letra mayúscula, constituye un conjunto de fincas con características comunes. Véase explicación en el texto.

- Las fincas situadas en el sector **A** estarán en las mejores condiciones: tendrán buena dotación de maquinaria y equipo, sus condiciones sanitarias serán apropiadas y el rebaño lechero será grande y productivo.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

Caracterización de la producción lechera de un distrito

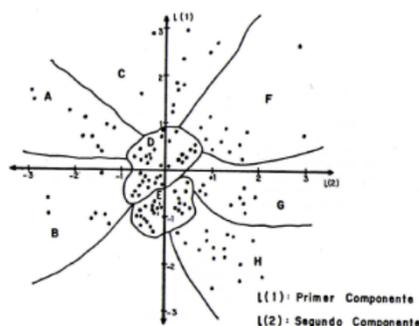


Fig. 5. Representación bivariada de resultados. Cada sector, identificado con una letra mayúscula, constituye un conjunto de fincas con características comunes. Véase explicación en el texto.

- Las fincas del sector **B** también serán satisfactorias en cuanto a equipo, maquinaria y condiciones sanitarias, pero la superficie total de la explotación no será tan grande ni el rebaño lechero tan numeroso, aunque los índices de productividad por vaca serán buenos.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

Caracterización de la producción lechera de un distrito

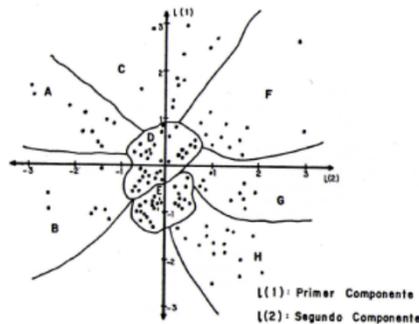


Fig. 5. Representación bivariada de resultados. Cada sector, identificado con una letra mayúscula, constituye un conjunto de fincas con características comunes. Véase explicación en el texto.

- El sector **C** comprenderá las fincas grandes con dotación heterogénea de recursos, pero posiblemente no muy productivas (segundo componente próximo a cero). Los sectores **F** y **G** estarán constituidos por fincas casi iguales, con superficie cercana al promedio, con cierto equipo, pero con baja productividad.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

Caracterización de la producción lechera de un distrito

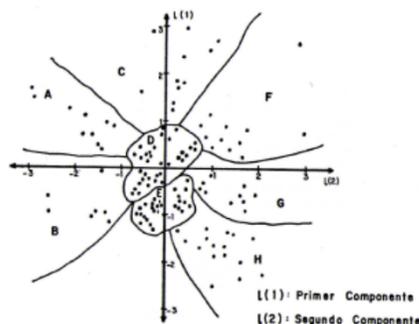


Fig. 5. Representación bivariada de resultados. Cada sector, identificado con una letra mayúscula, constituye un conjunto de fincas con características comunes. Véase explicación en el texto.

- Los sectores **D** y **E**, ubicados en el promedio de la zona, afrontarán diversos problemas, pero su productividad será cercana al promedio de la zona. El sector **D** contará con más recursos que el **E** para mejorar su productividad. Las fincas del sector **H** tendrán baja productividad debido a la falta de recursos para explotarlas.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

ACP - RESUMEN

- ✓ ACP se describe como un método para producir una combinación lineal de las variables para las cuales se dispone de datos. Las combinaciones lineales son las CP.
- ✓ La primera combinación lineal es el primer CP, que tiene la propiedad de tener la mayor varianza posible. El segundo CP tiene la propiedad de tener la mayor varianza posible sin estar correlacionado con el primer CP. Los restantes CP se definen de forma similar.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

ACP - RESUMEN

- ✓ Los CP se calculan encontrando los **valores propios y vectores propios de la matriz de covarianza** de la muestra para las variables X. Si las variables están estandarizadas (media cero y varianza unitaria), la matriz de covarianza es también la **matriz de correlación** para las variables X. En este caso, todas las variables tienen igual peso.
- ✓ Si los valores propios están en orden, desde el mayor al más pequeño, entonces **el i-ésimo valor propio es la varianza del i-ésimo CP, y el vector propio correspondiente da los coeficientes de las variables X para este PC.**

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

ACP - RESUMEN

- ✓ Si el análisis se lleva a cabo utilizando la **matriz de correlación**, entonces **la suma de los valores propios es igual a p** , el número de las variables X .
- ✓ Usualmente se consideran sólo los primeros CP, siempre que la suma de sus varianzas represente un porcentaje alto (por ejemplo, 80% o más) de la suma de las varianzas para los **p CP**. Alternativamente, si el análisis se lleva a cabo en la matriz de correlación, entonces habitualmente se utilizan los **CP con varianzas mayores que uno**, puesto que éstos tienen varianzas que son mayores que las varianzas de las variables estandarizadas X (es decir, representan más variación que cualquiera de las X variables originales).

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

ACP - RESUMEN

- ✓ **Modo R:** Dadas n observaciones sobre p variables se calcula la matriz de covarianza o correlación ($p \times p$) y se determinan los valores propios y vectores propios. Estos últimos se utilizan para representar cada vector de observaciones en términos de CP. En estos casos, **el análisis se centra en las variables (el objetivo es reducir el número de variables)**.
- ✓ **Modo Q.** Si el análisis se centra en las observaciones, se obtiene una matriz de covarianza o correlación ($n \times n$), y se determinan los valores y vectores propios. El objetivo del modo Q es **estudiar las interrelaciones entre observaciones y encontrar grupos de observaciones que sean similares entre sí en función de su composición total**.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.