55th EAAP Meeting **Bled, 5th – 9th Septembre 2004**

Session Pl5.5 **eliana.barros@portugalmail.pt**

The principal components analysis as a tool for a better management of bovines farming system.

J.C.-R. Santos and E.V. Barros, Divisão de Produção Animal (DRAEDM) – Rua Franca, n.º 534, 4800-875 S. Torcato, Portugal.*

Abstract

The principal components analysis method leads us to select a group of discriminating variables and reassembles the analysed farm based on its parameterisation. We studied a group of 35 farms where Minhota is the main local breed. This cow-calf breed is raised in the Northwest of Portugal, within a system where the main characteristics are its reduced number of animals per farm (1-6), permanent lodgement (83% are kept in the cattle-shed) and traditional feeding system (oat+rye-grass+rye: 1,7-16,9 Kg DM/cow.day; maize straw: 2,6-22,0 Kg DM/cow.day; maize silage: 1,1-21,0 Kg DM/cow.day), being the farms constituted by many patches (3-33) with reduced dimension (0,06-0,40 m^2). The farmers have an advanced age (31 % are above 65 years), low literacy degree (11 % cannot read and write) and low qualification level (66 % without technical courses). We choose 10 variables (from an initial group of 50) that present significant statistical correlations ($P<0,05$) and explain more than 50 % of the total variance. Based on these 10 variables, the ACP built the following equations that will allows us to reassemble the 35 farms according to its characteristics.

$Y_1= (0,487_{AMED}+0,423_{PAR}+0,415_{CABO}-0,093_{PNUM}-0,122_{AMAC}+0,168_{NAVE}+0,105_{1PAR}+$
$+0,396_{ISIL}-0,0921_{ÁRCA}+0,428_{AREJ})$ and

$Y_2= (0,234_{AMED}+0,29_{PAR}-0,146_{CABO}-0,173_{PNUM}+0,472_{AMAC}+0,531_{NAVE}-0,066_{1PAR}-$
$-0,248_{ISIL}+0,454_{ÁRCA}-0,181_{AREJ})$

Although it is necessary further studies, we can conclude that PCA can be used as a tool in order to decide the best strategy to improve bovine herd's management.

Key words: principal components analysis, bovines, livestock farming systems.

Introduction

The importance of the knowledge related to the working process of a certain livestock farming system should be echoed in the implementation of social and economical development strategies and/or of technical nature to him applied.

A livestock farming system is characterised by a great diversity of endogenous factors that limit his technical end results, but also by a multitude of exogenous factors that limit his social and economical end results. The livestock farming system understands two sub-systems. One, the decision sub-system, report to the system administration, guiding the production processes and techniques to the endings way. Other, the technological sub-system, reports to the tasks execution guaranteeing the raw materials into end products transformation functions (Gibon *et al.*, 1989).

For being a diversified (production of milk per cow: 1.016,8 (±353,47) Kg in 305 days; use of corn silage in the animal feeding: 66 % of the explorations) and powdered system (2 bovines heads/farmer, 2,6 (±1,86) hectares/farm, 17,6 (±8,06) patches/farm), the farming system of the local bovine race, called as Minhota, gathers a group of characteristics that allows us to rehearse the use of PCA as an analysis instrument, where we intend to rehearse the validation of the following statement: "The statistical method of PCA allows an analysis of any bovine livestock farming system in way that the variables that most discriminates the system's working process are clearly identified and quantified."

The Principal Components Analysis statistical method permit to transform a group of correlated variables and convert him into an independent variables matrix, in function of the initial ones, that explain the existent variance among the variables (Blasco, 1996; Silvestre, 2001).

Materials and methods

The bovines of Minhota's cattle breed are reared in triple aptitude, in other words, for its meat, milk and farm work. In this study were selected the farms where the animals were reared in this triple aptitude, for treating of a geographical area well delimited (parish of Correlhã in the Ponte de Lima municipality) and mainly because the farming of these animals for milk production is going to extinction. The method that we privileged to collect information relative to Minhota's milk production farming system was the closed answer inquiry. We have made 35 inquiries that correspond to 35 farms.

To characterise this livestock farming system we calculated the average and standard deviation as dispersion measures. The Principal Components Analysis was done through the JMP program, version 3.2.2. of SAS Institute Inc. (1989).

Results and Discussion

Of an initial lot of 50 variables were selected 10, because they present significant statistical correlations (P <0,05 (table 1)) and because they form a group of variables whose first two vectors explains more than 50 % of the total variance verified, being them: average farm dimension (AMED); number of patches per farm (PAR); total number of bovine heads (CABO); numeric productivity (PNUM); weaning age of calves for rear (AMAC); period between the calves' birth and sale (NAVE); age at first labour (1PAR); amount of corn silage supplied (ISIL); facilities area per bovine head (ÁRCA); air entrance area for natural ventilation (AREJ).

Table 1: Correlations matrix among the ten variables in study.

Variables	AMED	PAR	CABO	PNUM	AMAC	NAVE	1PAR	ISIL	ÁRCA	AREJ
AMED	1,00									
PAR	0,86	1,00								
CABO	0,62	0,59	1,00							
PNUM	-0,20	-0,04	0,05	1,00						
AMAC	-0,01	0,20	-0,17	0,30	1,00					
NAVE	0,44	0,41	-0,06	-0,49	0,24	1,00				
1PAR	0,10	0,09	-0,01	-0,03	0,03	0,20	1,00			
ISIL	0,56	0,36	0,41	-0,10	-0,49	0,03	0,31	1,00		
ÁRCA	0,12	0,04	-0,25	0,02	0,28	0,22	-0,36	-0,19	1,00	
AREJ	0,63	0,44	0,67	0,07	-0,25	0,05	-0,03	0,61	-0,21	1,00

The first principal component ($\hat{e}_1z$) explain 34,7 % of the total variance, while the two components ($\hat{e}_1z$, $\hat{e}_2z$) explains 54,2 % (table 2), for what we choose to use only these first two main components.

Table 2: Principal components ($\hat{e}_1$, $\hat{e}_2$, ..., $\hat{e}_{10}$) relative importance in the verified variance justification.

Principal Component	1st	2nd	3rd	4th	5th
Variance	3,47	1,94	1,51	1,22	0,74
% total variance explained	34,72	19,44	15,05	12,20	7,44
% accumulated	34,72	54,17	69,22	81,42	88,86
Principal Component	**6th**	**7th**	**8th**	**9th**	**10th**
Variance	0,42	0,28	0,22	0,14	0,06
% total variance explained	4,19	2,76	2,24	1,37	0,58
% accumulated	93,05	95,81	98,05	99,42	100,00

This way, and in agreement with the coefficients of the table 3, the following equations of the two referred vectors were built:

$$Y_1=\hat{e}_1 z=0{,}487_{AMED}+0{,}423_{PAR}+0{,}415_{CABO}-0{,}093_{PNUM}-0{,}122_{AMAC}+0{,}168_{NAVE}+0{,}105_{1PAR}+$$
$$+0{,}396_{ISIL}-0{,}0921_{ÁRCA}+0{,}428_{AREJ}$$

$$Y_2=\hat{e}_2 z=0{,}234_{AMED}+0{,}29_{PAR}-0{,}146_{CABO}-0{,}173_{PNUM}+0{,}472_{AMAC}+0{,}531_{NAVE}-0{,}066_{1PAR}-0{,}248_{ISIL}+$$
$$+0{,}454_{ÁRCA}-0{,}181_{AREJ}$$

Table 3: Principal components' vectors coefficient.

PC \ Var.	AMED	PAR	CABO	PNUM	AMAC
1st	0,487	0,423	0,415	-0,093	-0,122
2nd	0,234	0,291	-0,146	-0,173	0,472
PC \ Var.	**NAVE**	**1PAR**	**ISIL**	**ÁRCA**	**AREJ**
1st	0,168	0,105	0,396	-0,091	0,428
2nd	0,531	-0,066	-0,248	0,454	-0,181

The variables that play a part of larger relevance in the justification of the verified variance are the ones that, by the selected principal components graphic representation, are more distanced from the origin point (0,0) and closer to the axes (Blasco, 1996).

Letter: AMED- average farm dimension (ha); PAR- number of patches per farm; CABO- total number of bovine heads; PNUM- numeric productivity; AMAC- weaning age of calves for rear (months); NAVE- period between the calves' birth and sale (days); 1PAR- age at first labour (days); ISIL- amount of corn silage supplied (Kg/day); ÁRCA- facilities area per bovine head (m^2); AREJ- air entrance area for natural ventilation (m^2).

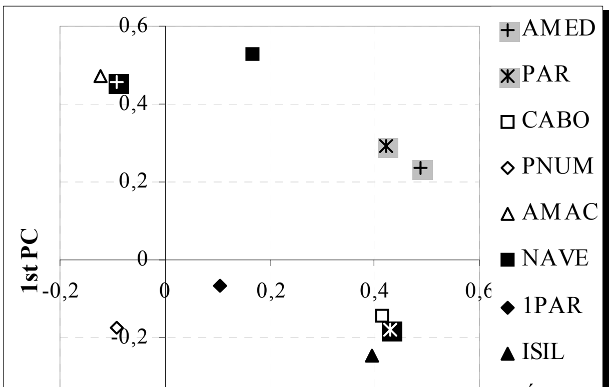

Illustration 1: Graphical projection of the ten selected variables on the 1st and 2nd principal components axis.

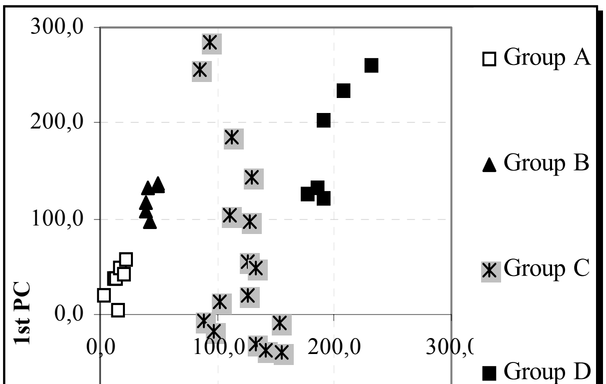

Illustration 2: Distribution of the farms on the principal components axis (Y_1 e Y_2).

In observation of the illustration 1, we don't find variables with a relevant role in the variance verified. However, the variables CABO and AREJ, for being closer of the 1st PC axis and, AMAC and ÁRCA, for being closer of the 2nd PC axis, are relatively more important than the other ones. As they form a 90º angle, the variables ÁRCA and AMAC explain different variation causes of the ones that are explained by the variables CABLE and AREJ (Blasco, 1996).

Rearranging the farms in function of the 10 selected variables, in agreement with the coefficient of each one of the variables present in the first two principal components (illustration 2), allowed us to build the table 4.

We consider the existence of four groups (A, B, C and D (illustration 2)). The more concentrated is the group more homogeneous are the farms that are part of it. This because, as the farms are distributed along the graphic in agreement with the values of the 10 selected variables, the more closed are the values of those variables and greater is the number of variables with close values less significant will be the variance between farms. The groups A (7 farms) and B (6 farms), for their concentrated graphic distributions, are formed by assemblages of farms identical amongst themselves, sharing, this way, of great similarity as much at the level of farms dimension and organization (AMED, PAR, ÁRCA, AREJ, CABO), as at the level of the productive and feeding practices (1PAR, AMAC, NAVE, ISIL) and even on the productivity resulting from those practices (PNUM). On the contrary, in the groups C (16 farms) and D (6 farms), on which the distribution is dispersed, the farms are not so identical amongst themselves.

Table 4: Values of the discriminating characteristics of the four groups.

Variables	Group A	Group B	Group C	Group D
AMED (ha)	1,4 ± 0,71	2,1 ± 1,53	2,4 ± 1,83	5,0 ± 1,21
PAR	11,3 ± 5,31	14,5 ± 7,50	15,9 ± 7,30	26,3 ± 5,24
CABO	1,3 ± 0,80	1,9 ± 0,92	2,5 ± 1,68	3,6 ± 1,89
PNUM	0,9 ± 0,09	0,7 ± 0,32	0,9 ± 0,21	0,6 ± 0,25
AMAC (months)	1,7 ± 0,81	2,0 ± 0,63	2,2 ± 0,83	2,7 ± 2,75
NAVE (days)	57,4 ± 26,86	209,7 ± 36,41	198,4 ± 151,36	438,7 ± 114,59
1PAR (days)	N/obs.	N/obs.	858,3 ± 209,26	953,1 ± 85,28
ISIL (Kg)	3,8 ± 7,19	4,0 ± 4,72	5,4 ± 4,35	7,3 ± 4,69
ÁRCA (m^2)	12,8 ± 6,53	9,5 ± 5,11	11,0 ± 7,12	9,2 ± 3,62
AREJ (m^2)	1,6 ± 2,82	0,0 ± 0,04	9,7 ± 25,31	15,1 ± 22,75

Letter: AMED- average farm dimension (ha); PAR- number of patches per farm; CABO- total number of bovine heads; PNUM- numeric productivity; AMAC- weaning age of calves for rear (months); NAVE- period between the calves' birth and sale (days); 1PAR- age at first labour (days); ISIL- amount of corn silage supplied (Kg/day); ÁRCA- facilities area per bovine head (m^2); AREJ- air entrance area for natural ventilation (m^2).

Based in the table 4 values we can define a strategy specifically directed to each one of the 4 groups found as a consequence of the PCA method application, with the purpose of lessening the strangling detected. In a general perspective, we point out the need to reduce the parameters PAR (number of patches per farm), as the farms present a high number of patches, AMAC (weaning age of calves for rear), since it is an agrarian system that seeks the milk commercialisation, and ÁRCA (facilities area per bovine head) once the verified values are the double of the foreseen for the biological agriculture (5 m^2/head (C.E., 1992)). In A and B groups we verified the need to increase the parameter CABO (total number of bovine heads), which would reflect in the decrease of the parameter ÁRCA (facilities area per bovine head), as well as the need to increase the parameter AREJ (air entrance area for natural ventilation) and ISIL (amount of corn silage supplied).

Using the great versatility of this method to accomplish sectorial or parcelled analyses we present following a concrete example of this possibility. Of the 50 variables that characterise the livestock farming system studied, we used the relative ones to the productive parameters to proceed to a sectorial analysis of the animal production of this farming system (age at the first labour - 1PAR; weaning age of calves for rear - AMAC; weaning age of calves for sale - AMAV; season of labours - ÉPAR; interval among labours - IPAR; lactation length - LACT; period between the calves' birth and sale - NAVE; milk production in 305 days - PLEI; numeric productivity - PNUM, and cows farm work - TRAB). From these 10 variables 5 were selected (AMAV, IPAR, NAVE, PLEI and PNUM) because these they present significant statistical correlations (P <0,05) and because they form a group of variables whose first two vectors explains more than 50 % of the total variance verified. The first PC ($\hat{e}_1 z$) explain 39,2 % of the total variance, while the first and second PC ($\hat{e}_1 z$, $\hat{e}_2 z$) explains 66,4 %. This way, the following equations regarding the two were built:

$$Y_1 = \hat{e}_1 z = -0{,}442_{PNUM}+0{,}488_{IPAR}+0{,}381_{AMAV}+0{,}485_{NAVE}-0{,}432_{PLEI}$$

$$Y_2 = \hat{e}_2 z = 0{,}581_{PNUM}-0{,}323_{IPAR}+0{,}568_{AMAV}-0{,}024_{NAVE}-0{,}485_{PLEI}$$

By the graphical representation of the two vectors (illustration 3) we verified that, for being distanced from the origin and on the 1st PC axis, the parameter NAVE (period between the calves' birth and sale) is the most important in the verified variance. It also can be said that the parameters PLEI (production of milk in 305 days) and AMAV (weaning age of calves for sale) have negative correlation, once their distribution is symmetrical.

Distributing the farms by the two vectors axes and in function of the 5 selected variables, in agreement with the coefficient of each one of the variables, we divided the studied farms in five groups: the group A (2 farms), the group B (8 farms), the group C (13 farms), the group D (10 farms) and the group E (2 farms).

Letter: PNUM- numeric productivity; IPAR- interval among labours (days); AMAV- weaning age of calves for sale (months); NAVE- period between the calves' birth and sale (days); PLEI- milk production in 305 days (Kg).

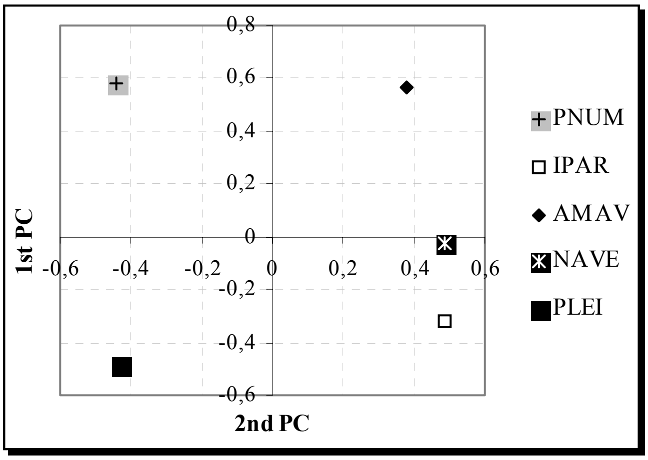

Illustration 3: Graphical projection of the ten selected variables on the 1st and 2nd principal components axis.

Formed the five groups and based in table 5 values become possible to draw directed strategies to each one of the groups, if necessary, with the purpose of to increase the productivity of the system and to reduce the difference verified among farms. This way, in order to make this system economically maintainable, it would be important to increase the parameter PLEI (milk production in 305 days) in all groups. In the groups B and C it will be necessary to reduce the parameter IPAR (interval among labours), which would reflect in an increase of the parameter PLEI. In the group D it could be increased the parameter PNUM (numeric productivity) through a reduction of IPAR. In the group E it also should be increased PNUM, similarly through an IPAR reduction. However, this procedure would not bring the same results for the group E that for C and D once the value of this parameter in the group E isn't as low as in the others.

Table 5: Values of the discriminating characteristics of the five groups.

Gr. \ Var.	PNUM	IPAR (days)	AMAV (months)	NAVE (days)	PLEI (Kg)
Group A	0,9 ± 0,18	365,5 ± 0,00	1,3 ± 1,06	21,3 ± 0,00	1776,4 ± 383,34
Group B	0,9 ± 0,20	410,9 ± 63,42	1,6 ± 0,78	143,3 ± 139,91	1347,3 ± 156,40
Group C	0,8 ± 0,20	429,2 ± 64,39	1,5 ± 0,75	176,7 ± 117,70	938,0 ± 156,62
Group D	0,7 ± 0,27	497,6 ± 133,11	2,1 ± 0,74	369,0 ± 152,20	702,9 ± 180,11
Group E	0,6 ± 0,34	383,4 ± 12,40	1,8 ± 0,35	127,7 ± 107,01	N/obs.

Letter: PNUM- numeric productivity; IPAR- interval among labours (days); AMAV- weaning age of calves for sale (months); NAVE- period between the calves' birth and sale (days); PLEI- milk production in 305 days (Kg).

After the collected information statistical treatment and the obtained results analysis, it is right to say that, through PCA, the discriminatory variables of the working process of this livestock farming system were identified and quantified, in the general analysis as well as in the factorial analysis. Based on the obtained results it was possible to outline appropriate intervention strategies to each one of the groups, for both analyses.

Also in a study accomplished by Santos (2002) in the Portuguese autochthonous bovine race Cachena, PCA was used in order to distinguish the variables that play a part of larger relevance in the verified variance justification. This way, from a 15 variables initial lot, seven were selected (total bovine heads - EFE; lactation length - DL; litters of milk milked in each lactation - LLL; period of free pasturing 24 on 24 hours - PP; age of the bulls at the first copula - COB; age in months at the first labour - PAR, and inquired literary qualifications - HL) because they are the ones that presented significant statistical correlations and the first two principal components explain 52,1 % of the verified variance. The obtained results were:

$$Y_1=\hat{e}_1z=0{,}372_{EFE}+0{,}555_{DL}+0{,}540_{LLL}+0{,}381_{PP}+0{,}240_{COB}+0{,}242_{PAR}-0{,}021_{HL}$$

$$Y_2=\hat{e}_2z=0{,}372_{EFE}-0{,}142_{DL}-0{,}119_{LLL}-0{,}043_{PP}+0{,}516_{COB}-0{,}369_{PAR}+0{,}650_{HL}$$

The most important variables in the justification of the variance are the ones related with the milk production (LLL and DL) once these are on the axis of 1st PC and the farmers' literary qualifications (HL) that is on the axis of 2nd PC. Among other, it was pointed by Santos (2002) the need to intervene in the instruction and professional specialisation sector, since in all of the groups in which the sample was divided (4 groups) the education level is very low (4th class on average and 6th class at the most). Also in this work PCA allowed to distinguish the most important variables in the justification of the variance verified among farms, as well as to quantify them. Equally, with the objective of using the information obtained in the direction of improving the efficiency of the livestock farming system some intervention strategies were drawn.

Already Filacorda *et al.* (2000) made another approach to PCA, applying it to determine the impact of an extensive livestock farming system (alpine pasture) in areas of Italian Alps nature reserve, finding out the relationship among the variables relative to the livestock presence (livestock, pastures areas and sites of community interest (SCI) stocking rate) and the ones relative to ecological values (number of *Malghe*, number of habitats, number of species, reptiles, mammals, birds, plants, invertebrate animals and SCI size), being that relationship explained by 2 factors. The animal production systems seemed to be positively enhance the presence of high habitats numbers, plants and birds; they verified a quick negative correlation among the farms, pasture areas and the livestock presence and the reptilian and invertebrates species; and the SCI stocking rate had a negative effect in the number of species, plants, habitats and mammals (i.e. carnivorous), particularly in smaller dimension areas. In this specific case, PCA had a relevant role in the determination of the most important variables of this livestock system and allowed to verify the dynamics between them. However, in the consulted bibliography wasn't found any intervention strategy, for what would be interesting to use the information obtained with the intent of to improve the performance of this bovine livestock farming system as well as to try to lessen the livestock effects in the ecosystem where it's in.

Conclusion

Sometimes, for a certain livestock system, are projected improvement strategies that have for base summarized information of a particular group that seems to be more or less homogeneous. A lot of times, due to no detected asymmetries in a general analysis of the information, the strategies are applied and don't have the expected effects. It is here that the advantage of the principal components analysis method remains. Through PCA the factors that cause larger variance are distinguished, being also determined the dynamics among them and, from these, assembled the elements that share a nucleus of common characteristics and separated elements different amongst themselves. Like this, it's possible to remove advantage of the profits and/or to act on weak points that would be camouflaged along with the whole initial information group.

Bibliography

- ANNICK, G., ROUX, M., VALLERAND, F., FLAMANT, J.C., 1989. Elements Conceptuels et Methodologiques pour l'Approche des Explotations d'Elevage: Quelques examples Français. In: 40éme Réunion Annuelle de la FEZ de systémes d'elevage, 27-31 Aôut 1989, Dublin, Irland.
- BLASCO, A., 1996. Analisis de Datos Experimentales Para Proyectos Fin de Carrera. Servicio de Publicaciones, Departamento de Ciencia Animal, Valência, 91 p.
- COMUNIDADE EUROPEIA, 1999. Regulamento Comunitário N.º 1804/ 1999 do Conselho de 19 de Julho de 1999.
- FILACORDA, S., FABRO, C., SUSMEL, P., 2000. Contribution of alpine pastures to special areas of nature conservation in the North Easter Alps. In: 50th EAAP Symposium on livestock farming systems, RAP Posieux, Switzerland.

- SANTOS, J., 2002. Cachena – Uma Raça de Bovinos (Primeira abordagem ao sistema de criação). DRAEDM, O Minho, a Terra e o Homem, 44: 35-44.
- SILVESTRE, A.L., 2001. Análise de Dados : Estatística Descritiva. Copyright@Editora Vulgata. Lisboa, pp. 299-310.