

Identifying homogeneous target groups of dairy producers: an application of combined principal component and cluster analysis

S. J. Staal, D. Njubi, and W. Thorpe

International Livestock Research Institute, POB 30709, Nairobi, Kenya

Introduction Research aimed at developing appropriate interventions to assist smallholder dairy producers requires a clear understanding of the overall systems environment. Smallholder dairy systems are shaped by production characteristics and technologies, but also by farm/household resource constraints and the market environment. For example, farms in poor agro-ecological zones yet close to urban centres may compete effectively for markets against distant farms located on better land, and their production strategies will reflect that market orientation. The appropriate characterisation of dairy producer farm/households should thus consider this multiplicity of factors, so that resulting research priorities will be more likely to address the factors to which producers themselves are responding, and ultimately to yield the desired impact. This paper describes the characterisation method used to identify patterns among smallholder dairy farm/households in Kiambu district in the highlands of central Kenya, and then from among them relatively homogeneous target groups of dairy producers. The approach used differs from previous methods by combining principal component and cluster analysis to allow incorporation of a greater range of variables.

Data The data were gathered during the first phase of a comprehensive collaborative dairy research and development project in the central Kenyan highlands. Information on the farm, household, and dairying activity was gathered from a stratified random sample of 365 households in July 1996, using a structured questionnaire.

Method The clustering method follows Gockowski and Baker (1996), and is motivated by the desire to reduce the number of variables used without omitting important information. Traditional clustering methods require the selection of a few important variables. As more variables are included, sensibly interpreting the cluster results grows more difficult. The use of fewer variables, however, increases the chance of not including important variation. The method used here alleviates this constraint by identifying the most important variation from a larger set of variables, which is then used to cluster the farm/household observations. The process thus consists 1) principal component analysis of sets of farm/household variables to identify, within their vector space, a smaller number of new vectors along which most of the variation is observed to occur, and 2) scoring of the farm/households along the new vectors, creating new variables which are then used in a standard cluster analysis. The principal components of a matrix of farm/household variables correspond in order to the largest characteristic roots of the covariance matrix. The vector elements of the principal components are simple correlation coefficients between the original variables and the principal component. These can be used to create individual household observations along a new variables by their linear combination with the original household variable values. The new variables so created are then used in a cluster analysis.

The groups of variables used were selected *apriori* on the basis of “themes” considered centrally important to the theoretical determinants of intensification. The themes chosen were 1) level of intensification of the dairying activity, 2) the farm/household resources available, and 3) the level of access to markets and services. A set of variables was chosen for each theme. Dairying intensity was postulated to be measured by the availability of feeds per animal, and so the variables chosen were acres of maize and Napier grass planted per unit of dairy cattle (TLU), land available per TLU, and level of fodder and concentrate purchased per TLU. The variables selected to reflect household resources were female-headedness, off-farm employment, household income level, and total land. Variables for market access were distance of the farm from Nairobi, the availability of co-operative AI services, the farm-gate price of milk, co-operative membership, and milk sales to informal market outlets. To obtain complete data for all the variables used, the number of dairy household observations was reduced to 172.

Table 1 Variable scores (correlation coefficients) on derived variables (principal components) for 3 sets of variables.

Theme 1: Intensification			Theme 2: House Hold Resources		Theme 3: Market Access	
Variables	Prin Comp 1 OWNFODD	Prin Comp 2 PURCFODD	Variables	Prin Comp 3 WEALTH	Variables	Prin Comp 4 COOPPART
Maize/TLU	0.69	0.10	Female-headed	-0.39	Dist. to Nairobi	0.68
Napier/TLU	0.80	-0.08	% off farm employ	0.47	Co-op AI available	0.65
Conc/TLU	0.62	0.07	Income level	0.69	Farm gate price milk	-0.59
Fodd/TLU	0.03	0.99	Total land	0.76	Co-op membership	0.77
Land/TLU	0.87	-0.01			Informal marketing	-0.75

Results Principal component analysis was carried out on each set of variables, yielding 4 to 5 factors in each. Gockowski and Doyle (1996) suggest that a rule of thumb for selecting significant principal components is to retain those with eigenvalues of greater than one. On that basis, two principal components were retained for intensification, one for household resources, and one for market access. Table 1 shows the correlation coefficients of the selected principal components with the original variables. The principal components for intensification (Table 1) reveal two distinctive patterns: the first factor is heavily weighted by on-farm feed resources, indicated by correlation with land, Napier and land. This factor defines a new variable, OWNFODD, which can be considered an index of the level of use of fodder produced on the farm. The second factor is almost perfectly correlated with purchase of fodder, and

represents a new variable PURCFODD, an axis of variation separate from the first. Under household resources, the factor identified is seen to be strongly correlated with income, total land, the % of adults working off-farm, and negatively correlated with female-headedness. This factor was thus viewed as an index of wealth. Finally, the market access variables yielded one factor which shows strong correlation with distance to Nairobi and co-operative membership, and negatively with participation in the informal market, and so was named COOPPART, an index of level of association with formal, co-operative-based input and output markets.

Cluster analysis was then carried out using the new variables, and 8 clusters of farm/household observations were created. Four of those clusters were found to contain 91% of the farm/household observations, and were retained. Table 2 shows values of simple cluster means for a number of farm/household variables. The largest, cluster 3, shows low levels of own fodder resources, purchased fodder and wealth but active co-operative participation, and so this group were labelled the Co-op Resource Poor (CRP). The second-largest, cluster 4, is similarly resource poor but exhibits low levels of co-operative participation, relying instead on informal marketing of milk, and so were labelled the Informal Resource Poor (IRP) group of dairy farmers. Cluster 5 farmers exhibited high levels of wealth, co-operative participation and off-farm employment, and were designated the Community Leaders (CL). Finally, cluster 6 farmers are distinguished primarily by the relatively large amounts of fodder they purchase. These were labelled the Specialists (S), as they appear to purchase feed in order to practice specialised dairying.

Table 2 Means of farm, household and market characteristics for 4 clusters and for the total survey sample.

Cluster label and number	Co-op res poor (3)	Informal res poor (4)	Community leaders (5)	Specialist (6)	Overall
Number of Households	56	48	36	16	172
Farm size (acres)	2.24	1.69	4.94	2.04	2.93
Napier acreage (acres)	0.38	0.27	0.82	0.28	0.50
Dairy cattle TLU	1.74	1.66	2.24	1.24	1.70
Napier acres per TLU	0.25	0.18	0.40	0.26	0.36
Fodder purchased (Ksh/TLU/yr)	342	998	386	3,769	1,145
Concentrate purchased (Ksh/TLU/yr)	5,552	3,585	6,457	8,165	6,273
Female-headed (%)	32	38	6	19	27
Adults working off-farm (% of adults)	21	29	47	25	31
Income category	2.01	2.67	4.08	2.75	2.8
Distance to Nairobi (kms)	40.0	24.4	43.4	38.2	36.9
Co-op membership (%)	88	10	94	63	65
Informal milk market participation (%)	30	98	31	56	51
Avg price milk (ksh/lt)	12.59	14.80	12.66	12.95	13.24

Discussion The analysis identifies two resource-poor groups of dairy farmers which constitute some 60% of Kiambu dairy farmers, but which are differentiated primarily by their level of participation in co-operatives and the informal milk market. The IRP group, however, receive a substantially better milk price due to better prices available in the informal market - they also average half the distance to Nairobi of the other groups, resulting in informal market access. Both of the resource poor groups are distinguished by high levels of female-headedness. Of note are the relatively high levels of purchased feeds seen in both groups, with the CRP buying higher levels of concentrates, while the IFR buying relatively more fodder. This is likely related to access to feed purchases on credit by the CRP through co-operative membership. Another observed pattern is the Specialist group, who spend far more on both concentrates and fodder than any other group, in spite of average income levels, and who tend to be male-headed. No other variables appear to distinguish them. Finally, the CL group clearly have more land and higher levels of income, and plant more Napier and maize per animal. This group is almost entirely made up of co-operative members, about half the adults work off-farm, and only 6% of households are female-headed. Of note are the relatively low levels of concentrate purchase, which are in the same range as the other groups. The analysis has identified differences in feeding strategy, land use, and market and institutional choices between otherwise apparently similar farmer groups. Important questions remain, however. Research can now be targeted at understanding these differences, particularly among the resource-poor groups, with a view to identifying interventions which can assist them to move to higher levels of productivity.

Gockowski, J. J. and Baker, D. 1996. Targeting Resource and Crop Management Research Through an Ecoregional Methodology for the Humid Forest of Central and West Africa. Paper presented at 1996 Biennial Meeting of Rockefeller Social Science Research Fellows, Aug. 15, 1996, Nairobi, Kenya

Staal, S.J., L. Chege, M. Kenyanjui, A. Kimari, B. Lukuyu, D. Njubi, M. Owango, J. Tanner, W. Thorpe and M. Wambugu, 1997. Characterisation of Dairy Systems Supplying the Nairobi Milk Market: A Pilot Survey in Kiambu District for the Identification of Target Groups of Producers. KARI/ILRI Collaborative Research Project Report.