Exploratory data analysis of the FAO/DCI dataset on "Improving livelihoods of HIV/AIDS affected households in Northern Province, Zambia"

Alessandra Garbero

Abstract

The aim of this paper is to apply exploratory techniques in order to uncover vulnerability dimensions that characterize clusters of households within the Zambian dataset.
Specifically, our purpose was to discover latent factors and to build a typology of households according to the classification results.
The relationship between poverty and AIDS impact is a strong one that is known already. What is less understood is the relative importance and dynamics of different morbidity profiles across wealth groups in relation to household food security. This paper identifies some possible relationships in this regard.
The datasets we used belong to a quantitative baseline survey that was conducted in Northern Province Zambia, by FAO in February/March 2004 and funded by Development Cooperation Ireland (DCI).
The complementary use of two techniques, multiple correspondence analysis and hybrid cluster analysis allowed us to derive a categorisation *a posteriori* which confirms and adds more information to the *a priori* taxonomies hypothesised in the research design of previous qualitative studies (FAO, 2004).
Our results underline the presence of an apparent gender divide: female headed households seem to clearly cluster with vulnerability factors (chronic morbidity; the burden of keeping orphans; economic vulnerability factors; vulnerability to food insecurity) as opposed to male headed households, that in most of the cases are married and also prosperous in terms of economic status.
Chronic morbidity is severely harming the capacity of certain types of households to work and cope with the already negative circumstances, making them progressively destitute and also socially excluded. We also discuss the implications for both applied research and practical development interventions of the use of such exploratory techniques to evaluate vulnerability taxonomies.

Key words: exploratory data analysis, multiple correspondence analysis, cluster analysis, HIV/AIDS, vulnerability**.**

Date: 15th of June, 2005

INTRODUCTION **3**

AIM OF THE PAPER **3**

OUTLINE OF THE METHODOLOGY **6**
CONSTRUCTION OF COMPOSITE INDICATORS: ASSET INDEX AND LIVESTOCK OWNERSHIP INDEX 6

MULTIPLE CORRESPONDENCE ANALYSIS (MCA) **7**

CHOICE OF ACTIVE VARIABLES 7
RESULTS 8
INTERPRETATION OF FACTORS 10
I FACTOR: AFFECTED VS. NON-AFFECTED 11
II FACTOR: FEMALE FARMERS AND BEER BREWERS NON-EDUCATED & NON AFFECTED VS. MALE EDUCATED 13

CORRESPONDENCE ANALYSIS AND CLUSTER ANALYSIS **15**

METHODOLOGICAL FUNDAMENTALS 15
OUTLINE OF THE MAIN RESULTS 16

CONCLUSION **21**

REFERENCES **22**
APPENDIX 1: CLUSTERS DESCRIPTION **24**

Introduction

HIV/AIDS has spread rapidly in the developing world for a complex set of reasons, which vary between countries and regions. Most cases are through sexual transmission.
Others involve mother-child transmission, unsafe injecting and blood transfusions. While risky sexual behaviour may be the direct cause, underlying factors drive risk, such as poverty, vulnerability and cultural norms.
Women are particularly vulnerable. They are biologically more susceptible to infection than men and less able to exercise control over their sexual lives. Violence against women contributes further to their vulnerability.

HIV/AIDS is increasingly adding to the burden of chronic poverty and destitution in Africa. People living in poverty are vulnerable. Those marginalized by society – including men who have sex with men, sex workers, injecting drug users and migrants – have been at most risk. They have less access to information, supplies such as condoms, reproductive health services and good quality public services.

Poverty continues to be an endemic problem in Zambia. According to the Living Conditions Monitoring Survey of 1998, 73% of the population were classified as living in poverty with poverty rates estimated at 83% for the rural population. While poverty has long existed in Zambia, it is clear that diseases, including HIV/AIDS, have exacerbated this poverty by contributing to decreased agricultural productivity and to increased household food insecurity.

In 2000, the Government of Zambia began the development of the Poverty Reduction Strategy Programme (PRSP), whose overall objective is poverty reduction and economic growth. The PRSP (2002 - 2004) identifies agriculture, tourism, and transport and energy infrastructure as key sectors for economic growth and HIV/AIDS as one of the priority areas for social investment. The Government of Zambia recognises that HIV/AIDS is seriously undermining all its efforts in reducing poverty. Development Cooperation Ireland (DCI) supports the implementation of the PRSP and has decided to re-engage in the economic sector in Northern Province. The main entry point for this re-engagement is the negative impact of HIV/AIDS on the household economy of the poor rural.

It is against this background that DCI, FAO and the Government of Zambia, through the Ministry of Agriculture and Cooperatives, conducted Household Livelihood Research in Northern Province in order to get a better understanding of pro-poor economic development issues for vulnerable and HIV/AIDS affected households. The research, lasted eight months, was guided by FAO Headquarters and a Project Steering Committee on HIV/AIDS and Rural Livelihoods, comprising five representatives from FAO-Zambia, Development Cooperation Ireland, Zambian Government, the Central Statistics Office and a Non-Governmental Organisation (FAO 2004).

Aim of the paper

The aim of this paper is to apply exploratory techniques in order to uncover vulnerability dimensions that characterize clusters of households within the Zambian dataset.
Vulnerability is a sufficiently general concept so as to encompass many dimensions of well-being. One could think of vulnerability in terms of the likelihood that a new-born will grow more slowly than anthropometric norms, the likelihood that an 18 year old woman will die during childbirth in the approximately 20 years she will be of childbearing years, the likelihood that a seven year old will complete primary school. Vulnerability can be assessed at the individual or household level; it can also be aggregated over these units of observation.

For instance in the economics literature, "a household is said to be vulnerable to future loss of welfare below socially accepted norms caused by risky events. The degree of vulnerability depends on the characteristics of the risk and the household's ability to respond to risk. Ability to respond to risk depends on household characteristics – notably their asset-base. The outcome is defined with respect to some benchmark—a socially accepted minimum reference level of welfare (e.g., a poverty line). Measurement of vulnerability will also depend on the time horizon: a household may be vulnerable to risks over the next month, year, etc" (Alwang J. et al. 2001). Assessments of vulnerability are particularly concerned with downside risks, those that cause welfare to fall.

In the context of HIV/AIDS, the concept of vulnerability has emerged instead within the livelihoods literature in a number of important conceptual frameworks worth mention.

Barnett and Blaikie (1992) building on work by Gillespie (1989) elaborated a concept of relative vulnerability to labour loss of different farming systems. Three major indicators, that could, in the main, be established for defined farming systems from secondary literature, at least in Uganda where the research was carried out, were used to construct an algorithm to classify farming systems on a scale from most vulnerable to not vulnerable. These indicators were:

• Whether the farming system already has a shortage in energy or protein,
• Whether labour supply falls short of labour demand at any time in the agricultural calendar, or is less than 20% in excess of labour demand,
• Whether there are substitutable staple crops requiring a lower level of maximum labour input, which will provide sufficient energy and protein.

They also identified four further indicators of vulnerability:

• Significant economies of scale in returns to labour involved in major agricultural operations,
• A marked sexual division of labour, which would accentuate the loss of labour for specific tasks,
• Significant essential maintenance costs of soil conservation,
• A significant degree of out-migration and non-agricultural income brought about by population pressure on land. (Barnett and Blaikie 1992)

Subsequent work by Barnett et al. (1995, see also FAO 1995) described the utility of such an approach and the opportunity it presents for early warning of downstream impacts of AIDS. It also highlighted the importance of the interface between domestic and farm labour demand (Barnett et al. 1995).

Loevinsohn and Gillespie (2003), synthesising the work of the RENEWAL network (Regional Network on HIV/AIDS, Rural Livelihoods and Food Security) presented the "HIV/AIDS Lens" as a "conceptual tool to help decision makers in agriculture and allied fields from farmers to policy-makers, to review situations and actions in the light of HIV/AIDS"(2002). What the lens is and its use in "assessment, analysis and action" is discussed further in Loevinsohn and Gillespie's paper, but essentially it is based on four concepts.

This fourfold structure derives from:

• A concern with both the "downstream effects" of AIDS on agricultural livelihoods and the way different livelihoods may hasten or slow the spread of HIV infection, and
• A desire to embed within the conceptual framework and the methodology the ways in which communities (and innovators within them) resist HIV/AIDS and its effects, as well as suffer them.

The four key concepts are therefore:

•Susceptibility: the chance of an individual becoming infected by HIV
•Resistance: the ability of an individual to avoid infection by HIV

•Vulnerability: the likelihood of significant impacts occurring at individual, household or community level
• Resilience: the active responses that enable people to avoid the worst impacts of AIDS at different levels or to recover faster to a level accepted as normal.
Loevinsohn and Gillespie (2003), under "vulnerability and resilience", discuss the impacts of AIDS, and responses to them, at levels ranging from the micro-biological (individual) through the "micro-environmental"(household) and "meso-environmental" (community/farming system) to the "macro-environmental" (regional or national). This schema places factors accelerating or retarding the development of clinical AIDS in HIV-positive people under the headings of vulnerability/resistance.
Barnett and Whiteside (1999, see also Barnett et al. 2000) conceptualize the boundary between susceptibility and vulnerability slightly differently as they pass over factors accelerating or retarding AIDS and limit the use of "vulnerability" to households and larger socio-economic units. Without dwelling on terminology, it seems useful to distinguish vulnerability/resilience factors at the individual/biomedical level from vulnerability/resilience factors affecting livelihoods and the wider economy. If we keep the susceptibility/resistance concept, this effectively produces a threefold schema: factors influencing the risk of exposure to HIV, factors influencing the development of AIDS, and factors influencing the impacts of AIDS on households and wider social units.
Recent FAO studies on the impact of HIV/AIDS on rural livelihoods (2003, 2004) conceptualize vulnerability to AIDS impacts using various definitions: presence of chronic illness, HIV/AIDS related mortality and presence of orphans, in the household.
Specifically, in the FAO 2003 case studies (Uganda, Namibia and Zambia Southern Province) vulnerable households were defined as "a household where at least one family member has been lost to HIV/AIDS, or HIV/AIDS related chronic illnesses (TB and pneumonia), or one in which at least one family member is suffering or suffered from frequent or long illness due to, or related to, HIV/AIDS during the recall period (1997-2002)". Households where no member has died of, or is suffering from, HIV/AIDS related diseases were classified as non-affected (or non-vulnerable).
In the Zambia, Southern province study, (FAO, 2003), due to respondents' unwillingness to report cases of chronic illness and HIV/AIDS related deaths, the definition of vulnerable households used a proxy indicator of the impact of the epidemic: i.e., caring for orphans, with orphans being defined as children up to 18 years old who had lost one or both parents through death.
In the FAO study in Northern Province, Zambia, (FAO, 2004), vulnerability was defined in terms of presence of chronic illness and presence of orphans in the household in order to analyze various impacts of HIV/AIDS on rural livelihoods.

Using the latter dataset[1] (the aforementioned quantitative baseline survey that was conducted in Northern Province Zambia funded by Development Cooperation Ireland – DCI), we decided to adopt an inductive approach, free from model based-assumptions and a priori categorisations.

We defined vulnerability according to various dimensions: economic vulnerability (measured through agricultural production and asset index); social vulnerability (the burden of keeping orphans); health vulnerability (chronic morbidity and presence of a PLWHA-person living

[1] The questionnaire, (for a description of the questionnaire modules see FAO 2004), was very comprehensive and allowed us to gather useful information on rural livelihoods in Northern Province, Zambia.

with AIDS); and vulnerability to food insecurity (through the number of meals given to adults and children).

Specifically, our purpose was to discover latent factors and also to build a typology of households according to the classification results.
We derived a categorisation a posteriori which confirms our qualitative theoretical assumptions (see FAO 2004).
Thus, this paper highlights the need to characterize and differentiate households according to the type of vulnerability dimension. The relationship between poverty and AIDS impact is a strong one that is known already. What is less understood is the relative importance and dynamics of different morbidity profiles across wealth groups in relation to household food security. This paper identifies some possible relationships in this regard.

This paper is a tentative response to the need for analysing the complex ways in which HIV/AIDS is affecting people's livelihoods and the impacts of livelihood insecurity on HIV/AIDS.

Outline of the methodology

Factor analysis (FA) has been applied, initially, in order to build composite indexes in the data matrix.
The analysis proceeded on two stages and methods like multiple correspondence analysis (MCA) and cluster have been adopted. Statistical Programmes such as SPSS (Statistical Package for Social Sciences, v. 11) and SPAD (Système Portable pour l'Analyse des données v.5) have been used.
SPSS was extensively used for data preparation and SPAD for advanced analysis.
The use of MCA leads to the creation of axes, where the initial variables are arranged according to coordinates obtained through the analysis. Through cluster analysis, modalities can be grouped according to common characteristics which are described now from the "synthetic variables" that have been produced from the MCA. Modalities are clustered by Ward's criterion, which is based on the least possible reduction in the variance. So the set of modalities breaks down into a number of clusters of modalities with common characteristics, described with the help of the initial variables.

Construction of composite indicators: asset index and livestock ownership index

In order to obtain a suitable data matrix for such advanced analysis, we went through a process of reshaping, merging files and aggregating variables.

We derived an asset index and a livestock ownership index, using the methodology applied by Filmer and Pritchett, (2001) in their work.
The authors tried to overcome and solve two questions: how to aggregate various asset ownership indicators into one variable to proxy for household wealth and how to attribute a system of weights to the items in question.
A number of solutions have been used in the literature to overcome the problem of choosing weights. First, equal weights of all the assets, which has as its only appeal not seeming as completely arbitrary as it really is. The second possible solution is to impose a set of weights. For instance, prices of various assets could be used to construct an index of household wealth, but this is possible only if the prices of various assets are available.
A third solution is to not construct an index but simply enter all asset variables individually in a multivariate regression equation. This is the approach recommended in Montgomery, Burke, Paredes and Zaidi (1997) for use in fertility or mortality regressions using DHS data. This approach does handle the problem of "controlling" for wealth in estimating the impact of non-

wealth variables. However, as recognized by Montgomery *et al* (1997), it does not identify the wealth effect as many assets play a both a direct and an indirect effect on outcomes. There is no way to infer from the unconstrained coefficients on the asset variables from a multivariate regression the impact of an increase in wealth. Hence, while in some sense a regression coefficients produce a linear "index" of the asset variables (that which best predicts the dependent variable) this "index" cannot be interpreted as the effect of an increase in wealth.

Thus the authors suggested to use the statistical procedure of principal component (which is closely related to factor analysis) to determine the weights for an index of the asset variables.
In our analysis, the single variables (wheel barrows, ox carts, grinding mills, axes, hoes, shovels, guns, cultivators, and bikes) have been used to create an index of assets that proxies for household "wealth" or economic status.
In the case of the livestock ownership index, we selected a number of significant assets such as other cattle, sheeps, goats, pigs and poultry and we formed a composite index, named livestock ownership index to be included as variable in the following analysis.
These two linear indexes of assets and livestock variables, proxies for household wealth and livestock ownership respectively, were employed as illustrative variables in the following multiple correspondence analysis.

Multiple correspondence analysis (MCA)

Multiple correspondence analysis is the multivariate version of simple correspondence analysis. Correspondence Analysis (CA) is primarily a technique for displaying the rows and the columns of a two-way contingency table as points in corresponding low-dimensional vector spaces. These spaces may be superimposed to obtain a joint display. With suitable care the analysis may be extended to display other matrices of non negative data.
CA is theoretically equivalent to a number of techniques which have appeared independently in the statistical literature since the mid 1930s, as simultaneous linear regression, reciprocal averaging, dual (or optimal) scaling. CA may be described as a further special case of discriminant analysis, called double discriminant analysis, where the dependence of two partitions of the same sampling units is investigated. Alternatively, the analysis may be considered as studying the dependence of two (or more) qualitative variables.

MCA is a less pure version but it is widely used in analyzing big questionnaires where the synthetic description made possible by the computer is unavoidable.
The starting point in MCA is a "particular" contingency table, named the Burt table, where the row modalities coincide with column ones. However:

- the general principles for interpretation are the same as those of CA;
- the percentage of inertia explained by the first axis is very low compared to that of the simple analysis (about 3 to 5%), but this measures gives here a very pessimistic idea of the share of inertia described by the graph (we will proceed to the re-evaluation of the inertias by applying Benzecri's formula, see following paragraphs).

Choice of active variables

In the analysis presented in this paper, there will be 11 active questions **Q** (and about 66 modalities of responses **J**). These questions are aimed at giving an overall description of

households characteristics (i.e. sex of household head, whether there is a person living with HIV/AIDS within the household i.e. PLWHA, whether the household is keeping orphans, marital status of head, most important and second important activity of head, age category of head, highest level of education attained, labour performance[2], nursing costs for illness and typology of illness suffered by the household head). The remaining variables have the status of illustrative variables and are variables of "performance" (i.e. asset index, livestock ownership index, comparison in area cultivated under maize between 1999-2004, comparison in quantity of maize harvested between 1999-2004, comparison in quantity of maize sold between 1999-2004), "proxies" for food security (number of meals had by adults and children per day) and two geographical variables (district and locality).

Table 1: Choice of the variables

```
SELECTION DES INDIVIDUS ET DES VARIABLES UTILES
VARIABLES NOMINALES ACTIVES
    11 VARIABLES      66 MODALITES ASSOCIEES
-------------------------------------------------------------------------------------------------
-----------------------------------
  11 . sex of head of household                                    (   2 MODALITES )
  13 . keeping person living with aids                             (   2 MODALITES )
  22 . keeping orphans?                                            (   2 MODALITES )
  31 . most important activity                                     (  15 MODALITES )
  32 . second most important activity                              (  15 MODALITES )
 840 . age category                                                (   4 MODALITES )
 860 . highest level of education attained                         (   5 MODALITES )
 865 . labour performance                                          (   5 MODALITES )
 869 . nursing costs for illness                                   (   4 MODALITES )
 890 . Disease                                                     (   7 MODALITES )
 891 . marital status                                              (   5 MODALITES )
-------------------------------------------------------------------------------------------------
-----------------------------------
VARIABLES NOMINALES ILLUSTRATIVES
     9 VARIABLES      44 MODALITES ASSOCIEES
-------------------------------------------------------------------------------------------------
-----------------------------------
 848 . asset index                                                 (   4 MODALITES )
 855 . livestock ownership                                         (   5 MODALITES )
 861 . locality                                                    (   9 MODALITES )
 863 . district (qual)                                             (   4 MODALITES )
 866 . maize area                                                  (   4 MODALITES )
 867 . maize harvested                                             (   4 MODALITES )
 868 . maize sold comparison                                       (   4 MODALITES )
 892 . adults meals                                                (   5 MODALITES )
 893 . children meals                                              (   5 MODALITES )
-------------------------------------------------------------------------------------------------
-----------------------------------
INDIVIDUS
---------------------------- NOMBRE -------------- POIDS ---------------
 POIDS DES INDIVIDUS: Poids des individus, uniforme egal a 1.                UNIF
 RETENUS ............ NITOT =    508      PITOT =             508.000
 ACTIFS ............. NIACT =    508      PIACT =             508.000
 SUPPLEMENTAIRES .... NISUP =      0      PISUP =               0.000
-------------------------------------------------------------------------
```

Results

As we said above we initially selected 11 active variables (**Q**) and 66 associated modalities of responses (**J**) and 9 illustrative variables with 44 associated modalities. MCA procedure eliminated all those modalities with zero or too weak weight (i.e. modalities with a frequency <2%) and randomly redistributed such frequencies into other modalities (of the same variable).

[2] Labour performance: how would you rate your ability to perform your duties (if chronically ill)?; nursing costs for illness: how were the nursing costs for illness covered?.

Thus, some modalities were involved in such process of re-distribution and a number of them were filtered out. The final number of active variables modalities decreased from 66 to 41.
As the common starting point of the MCA is an indicator matrix or a Burt matrix, correspondence analysis of the Burt matrix results in a diagonal matrix of principal inertias (or eigenvalues) and the matrix of standard coordinates (i.e. eigenvectors) . The number of nontrivial principal inertias is J-Q= 41-11=30 (see table 2). The inertias ("valeurs propre" in French), the percentages of inertias and the cumulative percentage of inertia are presented in Table 2.
The Trace (of the matrix) equals the sum of the eigenvalues (total inertia) and is equal to 2.7273.

Table 2: The values of inertias λk , the respective percentages of inertia and the cumulative percentages of inertia.

APUREMENT DES MODALITES ACTIVES
SEUIL (PCMIN) : 2.00 % POIDS: 10.16
AVANT APUREMENT : 11 QUESTIONS ACTIVES **66** MODALITES ASSOCIEES
APRES : 11 QUESTIONS ACTIVES **41** MODALITES ASSOCIEES
POIDS TOTAL DES INDIVIDUS ACTIFS : 508.00
VALEURS PROPRES
APERCU DE LA PRECISION DES CALCULS : TRACE AVANT DIAGONALISATION .. 2.7273
SOMME DES VALEURS PROPRES 2.7273
HISTOGRAMME DES 30 PREMIERES VALEURS PROPRES

NUMERO	VALEUR PROPRE	POURCENT.	POURCENT. CUMULE	
1	0.3646	13.37	13.37	**********
2	0.2008	7.36	20.73	**********
3	0.1624	5.96	26.69	**********
4	0.1556	5.71	32.39	**********
5	0.1360	4.99	37.38	**********
6	0.1237	4.54	41.91	**********
7	0.1177	4.31	46.23	**********
8	0.1112	4.08	50.31	**********
9	0.1072	3.93	54.24	**********
10	0.0977	3.58	57.82	**********
11	0.0954	3.50	61.32	**********
12	0.0917	3.36	64.68	**********
13	0.0896	3.28	67.96	**********
14	0.0881	3.23	71.19	**********
15	0.0859	3.15	74.34	**********
16	0.0814	2.99	77.33	**********
17	0.0780	2.86	80.19	**********
18	0.0723	2.65	82.84	**********
19	0.0712	2.61	85.45	**********
20	0.0684	2.51	87.96	**********
21	0.0623	2.28	90.24	**********
22	0.0593	2.18	92.42	**********
23	0.0550	2.02	94.43	**********
24	0.0437	1.60	96.04	**********
25	0.0430	1.58	97.61	**********
26	0.0331	1.22	98.83	********
27	0.0194	0.71	99.54	*****
28	0.0094	0.35	99.89	***
29	0.0026	0.10	99.98	*
30	0.0004	0.02	100.00	*

Because MCA includes fitting of the diagonal sub-matrices of the Burt matrix, the total inertia is inflated and thus the proportions of the first few principal inertias as part of the total inertia are reduced. One of the ways to address this problem, proposed by Benzecri (1979) is to consider only those principal axes, whose inertia are higher than 1/Q (i.e 1/Q=1/41= 0.024),

$$\tilde{\lambda}_k = \left[\frac{Q}{Q-1} \left(\tilde{\lambda}_K - \frac{1}{Q} \right) \right]^2$$

see Table 3.

Where k=1, 2,…..26
From the histogram of eigenvalues (Table 2) we can see that the first factor explains 13.4% of the total inertia, the second 7.36%, the third 5.96% and the fourth 5.71%. According to the histogram and the scree test (Table 2), it is wise to select the first 4 factors.

However let us calculate these inertias according to Benzecrí formula above.
We selected only those eigenvalues $\lambda_k > 1/Q = 1/41 = 0.024$ (which correspond to the first 26 factors) and we proceeded to the re-evaluation of such eigenvalues.

Table 3: The values of modified inertias (or eigenvalues) after applying Benzecrí's formula

Factor	λ_k	$\lambda_k - (1/41)$	$[\lambda_k - (1/41)]^2 = \lambda^*_k$	$\lambda^*_k / \Sigma\lambda^*_k$
1	0.3646	0.3402	0.1157	**42.45%**
2	0.2008	0.1764	0.0311	**11.41%**
3	0.1624	0.1380	0.0190	**6.98%**
4	0.1556	0.1312	0.0172	**6.31%**
5	0.136	0.1116	0.0125	**…**
6	0.1237	0.0993	0.0099	**…**
7	0.1177	0.0933	0.0087	**…**
8	0.1112	0.0868	0.0075	**…**
9	0.1072	0.0828	0.0069	**…**
10	0.0977	0.0733	0.0054	**…**
11	0.0954	0.0710	0.0050	**…**
12	0.0917	0.0673	0.0045	**…**
13	0.0896	0.0652	0.0043	**…**
14	0.0881	0.0637	0.0041	**…**
15	0.0859	0.0615	0.0038	**…**
16	0.0814	0.0570	0.0033	**…**
17	0.078	0.0536	0.0029	**…**
18	0.0723	0.0479	0.0023	**…**
19	0.0712	0.0468	0.0022	**…**
20	0.0684	0.0440	0.0019	**…**
21	0.0623	0.0379	0.0014	**…**
22	0.0593	0.0349	0.0012	**…**
23	0.055	0.0306	0.0009	**…**
24	0.0473	0.0229	0.0005	**…**
25	0.043	0.0186	0.0003	**…**
26	0.0331	0.0087	0.0001	**…**
			$\Sigma\lambda^*_k = 0.2727$	

Table 3 shows the values of modified inertias after applying Benzecrí's formula.
As we can see from table 3, the first four factors that initially explained 14%, 7% , 6% and 6% respectively, for a total of 27% of the total variance, now explain 42%, 12%, 7% and 6% , with a total of 67% of total variance. Therefore we decided to consider, only the first 2 factors.

Interpretation of Factors

In order to correctly interpret the factors obtained by means of MCA, we considered two criteria or approaches.
The first approach, named factorialist, is founded on the careful analysis of the absolute contributions of the variables and modalities of responses, together with their masses and index of distortion; at the same time we employed a second approach, named geometric–structural, more focused on the careful analysis of the shape of the cloud of points-modalities and on the distances among such points. The latter approach is founded on the analysis of the highest squared cosines, indicators of those points which are best represented on the axes.

I Factor: Affected vs. Non-affected

Table 4: Variables that give more contribution to the 1st factor

Variables	Cumulative absolute contributions	Cumulative %
Sex of household head	10.8	10.8
Keeping persons living with AIDS	7.6	18.4
Keeping orphans	3.6	22
Labour performance	17.7	39.7
How were the nursing costs for illness covered	18.2	57.9
Diseases	17.9	75.8
Marital status	12.9	88.7

If we examine the modalities of responses (of the variables presented in the table 4), their absolute contributions and the quality of their representation on the first axis (squared cosines), taking into account the sign of the factor coordinates, we obtain:

Table 5: Absolute contributions and squared cosines of modalities of responses that contribute to the I factor

Positive semi axis (+)	**Absolute contributions**	**Squared cosines**
Sex=male	3.9	0.43
Keeping PLWHA= no	2.6	0.30
Labour performance= normal	3.9	0.71
Nursing costs for illness= no	4.1	0.72
Disease= no chronic illness	4.1	0.71
Marital status= married	4.2	0.48
Negative semi axis (-)		
Sex= female	6.9	0.43
Keeping PLWHA=yes	5.0	0.30
Labour performance=below usual ability	9.8	0.46
Nursing costs for illness=sale of livestock, household goods and borrowing from relatives	8.1	0.37
Disease= HIV/AIDS-STDs	5.8	0.26
Marital status= widowed	7.7	0.40

This first factor seems to be characterized by the clear contrast between two poles, the affected vs. the non-affected in terms of morbidity status. Specifically, on one side (the positive semi axis "the non-affected"), we can find male headed households, those who are married , households which do not keep any person living with AIDS, whose labour performance is normal and that did not state to have any chronic illness.
On the opposite side, the affected ones with the following characteristics: mainly female headed households, keeping persons living with AIDS, whose labour performance is below usual ability and that are selling livestock, household goods and asking loans from relatives to cope with nursing costs for illness. Widow headed households fall within this pole and the prevalent chronic illness within such pole is HIV/AIDS.

If we observe the illustrative variables that belong to the first factor, we can see on the positive side, which is constituted by the non-affected, a medium-high asset index, a small livestock ownership index and in terms of food security, adults and children were given more than 3 meals every day. On the negative side, we notice those with a low asset index, that have a medium-high livestock ownership index and where adults and children were given only one meal.

The first factor marks a clear separation: on one side we find those who not only are vulnerable from a health perspective (i.e. are affected in terms of morbidity status) but also from a wealth perspective (low asset index). In addition, such factors are coupled with food insecurity, i.e. a not enough food intake, another vulnerability aspect that characterizes the "affected group".

On the other end we found nuclear households that are not affected by the fore mentioned vulnerability factors and are wealthier and healthier (see Figure 1).

II Factor: Female farmers and beer brewers non-educated & non affected vs. male educated

Table 6: Variables that contribute to the II Factor construction

Variables	Cumulative absolute contribution	Cumulative %
Sex of head of household	10.5	10.5
Most important activity	12.3	22.3
Second most important activity	14.2	36.5
Age category	3.0	35.5
Educational attainment	13.3	52.8
Labour performance	9.6	62.4
Nursing costs for illness	11.4	73.8
Disease	11.3	85.1
Marital status	8.9	94

Table 7: Absolute contributions and squared cosines of the modalities that contribute to the II Factor construction

Positive semi axis (+)	Absolute contributions	Squared cosines
Sex= male	3.8	0.23
Most important activity= formal employment	7.5	0.18
Second most important activity= farming	8.4	0.22
Age category=15-49	0.9	0.06
Educational attainment=post-secondary	5.4	0.12
Labour performance=below usual ability	5.1	0.13
Nursing costs for illness=households savings	5.0	0.11
Disease= HIV/AIDS-STDs	3.5	0.08
Marital status=married	3.1	0.20
Negative semi axis (-)	**Absolute contributions**	**Squared cosines**
Sex= female	6.7	0.23
Most important activity= farming	2.2	0.24
Second most important activity= beer brewing	2.7	0.06
Age category= 50-64	1.6	0.04
Educational attainment= no education	6.2	0.16
Labour performance= not stated	2.1	0.21
Nursing costs for illness=not stated	2.3	0.22
Disease= no chronic illness	2.3	0.22
Marital status= widowed	4.0	0.12

The second factor, which may result as confounding at first glance, is also very interesting, as another polarization is present. The variable that structures this separation is “gender”in association with economic and performance variables.

On the positive semi axis we can find modalities such as male headed household, married, formal employment, whose educational attainment is post-secondary and whose second most important activity is farming. They belong to the broad age group of 15-49 years old.
We also have modalities such chronically illness with HIV/AIDS-STDs as the prevalent reported diseases. The labour performance associated with such impairment is below usual ability and the release of households savings is the related modality of response to pay their nursing costs for illness.
On the negative semi axis, female-headed households are associated with categories such as farming as the most important activity and beer brewing as the second most important livelihood strategy. These are mainly widows aged 50 to 64 who have no education and did not state any chronic illness.
The analysis of illustrative variables provides other additional information about the typology of the populations which refer to the profiles characterizing the positive and negative semi axes: on the positive semi-axis, where male headed households are, we found that more than three meals were given to adults and children. Such typology also has a high asset index, but the latter lies on the border of significance threshold. Also such typology did not grow maize in both years (1999 and 2004).
On the other side, we found that that female headed households are associated with modalities such as less area cultivated under maize, less quantity of maize harvested compared with five years ago and a low asset index. One of the localities where such households are located and slightly significant is Kampumbu.
Figure 1 can be helpful in the interpretation and the visualization of the modalities in the bi-dimensional space.

Correspondence analysis and cluster analysis

Correspondence analysis rarely provides an exhaustive insight of a set of data. When processing the large data arrays issued form sample surveys through multiple correspondence analysis it often occurs that the results are still too complex to be easily read: the configurations of points need further summarizing. In many cases this can be achieved satisfactorily with the help of clustering techniques. The contribution of these techniques however is not limited to this practical phase of processing. The complementarity between correspondence analysis and classification concerns both the basic comprehension of data structures and the facilities provided in the final steps of interpretation of results. The viewpoints of the two approaches as well as their output are fundamentally different. Consequently a combined used of both techniques is highly recommended for a thorough description of any complex data set.
Correspondence analysis describes the main features of the data as they appear in the space spanned by the first principal dimensions. This could involve a substantial shrinkage (as a consequence of a projection onto a subspace) and/or some distortions due to the sensitivity to outliers of the principal axes.
A remote profile point can notably influence for example, the first principal axis, and therefore all the subsequent dimensions, since these dimensions are related to the first axis through the constraints of orthogonality of the axes. By contrast, most of the classification algorithms and particularly the agglomerative algorithms are locally robust in the sense that the lower parts of the produced dendrograms are largely independent of possible outliers. For all these reasons, we decided to complement our multiple correspondence analysis with a classification performed in the high-dimensional space spanned by all the significant axes[3].

Methodological fundamentals

The results of the typological study (obtained from a combination of MCA and clustering techniques) make sense only if the list of active variables is clearly specified.
These active variables are used to compute the distances or similarities between the respondents. If this list is sufficiently large, we can reasonably expect that the results obtained will be somewhat independent of the presence or absence of a particular variable within this set. The set of active variables must satisfy a criterion of homogeneity whose rationale is

---

[3] The two complementary techniques used to describe the DCI dataset are on the one hand multiple correspondence analysis, and the other hand a composite technique designated here as hybrid clustering. Since MCA has already been presented in other parts of the paper, we will give a basic idea about the clustering technique.
The technique of hybrid clustering comprises three steps:
1) Preliminary clustering, using agglomeration around variable centers (k-means method, or dynamic cluster method). The search for stable groups enables improvement and validation of the clusters. This search consists of obtaining several partitions, starting with different sets of provisional centers. The stable groups are the groups of individuals always clustered together (i.e. always belonging to the same cluster). In our example (see below), the 508 household heads were assigned to about 4 and 8 clusters.
2) Hierarchical clustering of the previously obtained clusters (using Ward's criterion), and determination of the cutting of the dendrogram, which determines simultaneously the number of final clusters, and a provisional set of centers for the final partition. The dendrogram provides unvaluable assistance in determining the number of clusters, a number which is unknown before hand. Inspection of the sequence of indexes suggests the most suitable partition which should correspond to a significant jump of the index.
3) Reallocation (using an iterative procedure similar to the k-means method) of the individuals or objects in order to improve the quality of the partitioning (even a relatively clear cut partition according to the shape of the dendrogram does not always produce a locally optimal partition).
Details about these methods and the corresponding software (SPAD) can be found in Morineau and Lebart (1986).

intuitive: the computations of the distances must make sense, so that patterns or groupings formed out of these distances also make sense. Whereas the homogeneous set of active variables allows for the definition of a specific point of view, the set of supplementary (or illustrative) variables will allow for a posteriori characterization or identification of the structural features produced. This set is not necessarily homogenous, since its elements are neither used to determine the basic MCA, nor the basic partition.
Our objective is to derive a typology (i. e. a set of households' classes) on the basis of the obtained partitions. The interpretation of the results aims at evaluating the classes, the whole partition, i.e. the derived typology.

Outline of the main results

MCA produce planar maps where the points represent the response-items and the proximities between points represent the affinities between these responses (i.e. two close responses have been given by almost the same individuals). The clustering will highlight the main grouping of individuals with respect to their most significant profiles. These groupings will be described in a systematic way, using all the objective characteristics of the individuals. The centroid of these groupings can be plotted onto the maps issued from MCA, thus providing assistance in deciphering and interpreting these rather complex outputs.

We performed cluster analysis to the DCI dataset and we obtained 2 partitions: 4 and 8 clusters. We decided to select the partition with 8 best clusters according to significance levels.

All the items characterizing the clusters are selected by the computer according to their value (or their percentage) within the cluster, as compared to their value (or percentage) in the global population (see Morineau and Lebart 1986).
The tables in appendix 1 show the lists of the modalities that contributed to each cluster. The significant modalities of responses were selected by means of the "valeur test" that needs to be greater than 2 (absolute value) to be significant at 95%.

We therefore proceed to the description of the 8 clusters obtained.
The final step will be to see how such conclusive 8 classes are represented in the 2 factors dimensional space.

Cluster I "Non-vulnerable households": 211 cases, 41.54% of the total sample.

This cluster includes households that do not present any of the vulnerability factors previously defined, such as economic and food insecurity, chronic morbidity, caring for a person living with HIV/AIDS and keeping orphans.
They present the following characteristics: male headed, married, whose main livelihood strategy is farming and whose second livelihood is fishing or none. The highest level of education attained is primary school.
Indicators of performance show that this group is rather wealthy in terms of : asset index (medium-high asset index 57.5% exclusive of the cluster; high asset index, 51.56 % exclusive of the cluster); and agricultural production (more maize harvested compared with 5 years ago, 57.97 % exclusive of the cluster). As far as the livestock ownership index is concerned, the latter is bordering significance level, with a "small/low" score, meaning that such households do not resort to livestock rearing as other clusters' members.

Cluster II "Fishing non-vulnerable households": 46 cases, 9% of the total sample.

This second cluster consists of households whose main livelihood strategy is fishing, trading (to a lesser extent) and whose second main activity is farming. Localities included in such cluster are Mumba and Muchinshi, and Mungwi and Chilubi district. Such households are mainly married and male headed. There are no chronic illness and do not keep PLWHA.
As far as agricultural production is concerned such households did not report any indication on the comparison variables, probably due to the fact that they are not growing any maize.

Cluster III "Formally employed non-vulnerable households": 36 cases, 7.09% of the total sample.

The third cluster is characterized by households who are formally employed and whose level of educational attainment is post-secondary. The second main livelihood strategy is farming.
As far as food security indicators are concerned, such households performed well, since adults and children are given more than 3 meals per day.

Cluster IV "Young women divorced/separated non-vulnerable": 44 cases, 8.66% of the total sample.
The forth cluster includes female-headed households, divorced and separated, falling in the 15-49 age group. They are characterized by beer brewing as their main second source of livelihood. We assume that they are practicing some farming, although it does not figure as modality in the output. We are able to derive such information as such group stated to have less area cultivated under maize, compared with five years ago.
They are not affected in terms of chronic morbidity since they did not report to have any illness.

Cluster V "Widows keeping orphans non-affected by chronic morbidity": 63 cases, 12.40% of the total sample.
This cluster is formed by female headed households, illiterate, specifically widows who are keeping orphans. They are not affected by any chronic illness, and they are mainly beer brewers and to a lesser extent farmers. They also stated household chores as secondary activity.
The presence of a low asset index score, as a proxy for household welfare, means that they should be relatively poor and thus economically vulnerable.

Cluster VI "Elderly-headed households over 65 affected by old age":27 cases, 5.31% of the sample.
This cluster is constituted by elderly-headed households over 65, affected by old age negative effects, diabetes and hypertension. Such group stated to be unable to work and to have a labour performance below usual ability, meaning that they are severely harmed by the associated chronic illness. They are selling livestock items, household goods and are borrowing money from relatives to meet nursing costs. They declared not to be engaged in any secondary activity. The only indicator of performance that figured in the list is a medium-high livestock ownership index which is consistent with the fact that such households are surviving from livestock rearing.

Cluster VII "Vulnerable households affected by fever-malaria": 24 cases, 4.72% of the total sample.
This cluster seems to be of residual nature. Its members are indeed vulnerable households, in terms of morbidity status, that do have a chronic illness, prevalently fever and malaria and to

a less extent HIV/AIDS (bordering levels of significance). They are mainly giving away household savings to pay for nursing costs, and selling crops too (to a minor extent). They brew beer as second livelihood strategy. Their labour performance is below usual ability. Some of such households seem to be widow-headed.

Cluster VIII "HIV/AIDS (and related chronic illness) affected households": 57 cases, 11.22% of the total sample

This cluster includes all the HIV/AIDS affected households with a labour performance below usual ability. Primarily they are selling livestock items, household goods, and asking loans from relatives to pay for their nursing fees. They are also selling crops to a less extent. This cluster presents also those households affected by pneumonia, tuberculosis and skin rash, and those who stated to have a unknown chronic illness. There is a prevalence of female headed households and some widows (also divorced and separated, although the latter is not typical of the cluster). In addition, some of such households said to be unable to work.

This cluster results highly food insecure as adults were given only one meal per day.

The only locality that appears in the list is Lukulu and the district is Mpika (bordering significance).[4]

[4] We need to bear in mind that since locality and district were not significant in our bivariate analysis, their closeness to the centroid of the first and second factor, means that such variables are non associated (therefore are independent).

Figure 1: 1st and 2nd factors dimensional space.

As we can see from Figure 1, Cluster I, II and III, respectively “Non-vulnerable households”, “Fishing non-vulnerable households”, “Formally employed non-vulnerable households”, are well represented on the positive side of the first factor “the non-affected”. Cluster V “Widows keeping orphans non-affected by chronic morbidity” is rather close to the centroid, meaning a weak degree of dependence when observing the affected-non affected factor space.

Cluster VI “Elderly-headed households over 65 affected by old age”, VII “Vulnerable households affected by fever-malaria” and VIII “HIV/AIDS (and related chronic illness) affected households” are strongly positioned on the opposite side among “the affected”.

Cluster V (“Widows keeping orphans non-affected by chronic morbidity”) has a very high “valeur test” (significance level) on the 2nd factor (negative side). This is coherent with the polarization that we found on the second factor (which we labelled “Female farmers and beer brewers non-educated and non-affected vs. male educated”). Cluster V perfectly meets the definition of the 2nd factor. Moreover, Cluster V is opposed to cluster III “Formally employed non-vulnerable households” on such factor.

Conclusion

Multivariate descriptive techniques are generally intended to discover something and not to prove anything.
Some limitations in the results lie in the nature of the data itself. For instance, "the personal level" i.e. the fact that most of the variables are based on self–assessed measures, or subjective judgements, could be responsible for structural features through the effect of latent factors.

Concerning the combined use of correspondence analysis and clustering, four points need to be stressed, which will serve as a conclusion of the paper:

1. *the descriptive power of the clusters*: it is easy to describe a group. It suffices to compute for category or modality of the nominal variable, the percentage of respondents within the group and the global percentage. The selected test of significance (valeur test) allowed us to select and sort the most characteristic items.
2. *the supplementary information provided by the clusters*: the clusters are not only used to describe regions of the factorial plane. They are established in a high dimensional space, supplying many more elements of information, and in particular information that could be hidden by the projection onto a two and or three dimensional sub-space.
3. *the descriptive power of the axes, the important of latent factors and consequently the importance of the spatial configuration of points*: this contribution of MCA is apparent in Figure 1; for instance the opposition of affected vs. non affected households is extremely important since we were able to locate the differentiated 8 typologies, issued by the partition.
4. *the discovery of latent phenomena*: independently of the "classification effect" of the clustering (used to divide a huge space in smaller portions easier to analyse), this technique, can evidently help to discover possible existing groups. Similarly correspondence analysis can put forward some unexpected latent factors. To uncover such hidden patterns, or dimensions, is indeed the original goal of both methods, and the most ambitious one. Their complementary use is necessary to achieve this quest as it is necessary to undertake the more modest task of description.

The complementary use of the two techniques has led us to a rich description of the DCI dataset thanks to the obtained typologies.
In the 8 clusters classification we found an apparent gender divide: female headed households seem to clearly cluster with vulnerability factors (chronic morbidity; the burden of keeping orphans; economic vulnerability factors; vulnerability to food insecurity) as opposed to male headed households, that in most of the cases are married and also prosperous in terms of economic status.
In addition, marital status is also another variable that structures the gender divide: men tend to remarry and therefore there is a very high prevalence of widowed and divorced women, which in none of the clusters appeared to be as wealthy as their married counterparts.
Besides evident gender differentials, particularly vulnerable appear to be elderly headed households, especially the ones who are chronically ill and over 65, unable to work.
Therefore we can conclude that a clear cut structure is apparent in our 8 clusters partition. The first 5 clusters where morbidity is absent: married farmers, fishermen, formally employed heads, single young women brewing beer and healthy widows keeping orphans. On the other side the last three clusters: chronically ill elderly headed, fever and malaria

affected households and HIV/AIDS and its related opportunistic diseases affected households (with a prevalence of female headed households).
It seems that chronic morbidity is severely harming the capacity of such households to work and cope with the already negative circumstances, making them progressively destitute and also socially excluded.

Such clusters structure confirms and add more information to the *a priori* categorisation that was hypothesised in the qualitative studies research design (see FAO, 2004). The vulnerability categories, as resulting from the DCI research hypothesis, aimed at differentiating households according to their exposure to different vulnerability factors.

In this essentially methodological paper, exploratory data analysis (EDA) achieved its scope of describing the DCI sample in depth, but in addition it provided evidence that our *a posteriori* categorisation is very close to the *a priori* and research design-based vulnerability categories.

References

ALWANG J, SIEGEL P B, AND JORGENSEN S L, 2001 "Vulnerability: a view from different disciplines" World Bank

BARNETT T AND P BLAIKIE 1992, *AIDS in Africa: Its Present and Future Impact*, Belhaven Press, London

BARNETT T, J TUMUSHABE, G BANTEBYA, R SSEBULIBA, J NGASONGWA, D KAPINGA, M NDELIKE, M DRINKWATER, G MITTI AND M HASLWIMMER 1995, "The Social and Economic Impact of HIV/AIDS on Farming Systems and Livelihoods in Rural Africa: Some Experience and Lessons from Uganda, Tanzania and Zambia", *Journal of International Development* **7**(1):163-176

BENZECRI, J.-P. 1980, *L'analyse des données tome 2 : l'analyse des correspondances*, Paris : Bordas

BENZECRI, J.-P. 1992, *Correspondence Analysis Handbook*, New-York : Dekker, P. 1994, L'analyse factorielle, Paris : PUF

BOLASCO, S. 1999, *L'analisi multidimensionale dei dati*. Rome, Carocci ed.

CIBOIS, P.,. 1994, *L'analyse factorielle*, Paris : PUF

FAO 1995, *The Effects of HIV/AIDS on Farming Systems in East Africa* FAO, Rome

FAO 2003, *HIV/AIDS and Agriculture: Impacts and Responses. Case studies from Namibia, Uganda and Zambia*. FAO, Rome. Downloadable at: http://www.fao.or/sd/2003/PE12013_en.htm

FAO 2004. *HIV/AIDS, gender inequality and rural livelihoods. The impact of HIV/AIDS on Rural Livelihoods in Northern Province, Zambia*. FAO, Rome.

FILMER, D., PRITCHETT, L. 2001. *Estimating wealth effects without expenditure data-or tears: an application to educational enrolments in States of India*, Demography, Volume 38-Number 1.Pag 115-132.

GODDARD, J., KIRBY, A., 1976, *Introduction to Factor Analysis*, London : Catmog, M.J.

GREENACRE, M.J., 1984, *Correspondence Analysis*, London : Academic Press

LEBART, L., MORINEAU, A., AND PIRON, J. (1995**).** *Statistique Exploratoire Multidimensionnelle*, Dunod, Paris.

LOEVINSOHN M. AND S. GILLESPIE 2003, "HIV/AIDS, Food Security and Rural Livelihoods: Understanding and Responding" RENEWAL Working Paper No.2, IFPRI/ISNAR, Washington and the Hague, available at : http://www.isnar.cgiar.org/renewal/pdf/RENEWALWP2.pd

Montgomery, M., Burke K., Paredes E., and Zaidi S., (1997)."Measuring Living Standards with DHS Data: Any Reason to Worry?" mimeo, Research Division, The Population Council. New York, NY.

TABACHNIK, B.G. 1989, *Using Multivariate Statistics*, New-York : HarperCollins

Appendix 1: Clusters description

DESCRIPTION DE LA Coupure 'b' de l'arbre en 8 classes
CARACTERISATION DES CLASSES PAR LES MODALITES
CARACTERISATION PAR LES MODALITES DES CLASSES OU MODALITES
DE Coupure 'b' de l'arbre en 8 classes

CLASSE 1 / 8

V.TEST	PROBA	---- POURCENTAGES ---- CLA/MOD	MOD/CLA	GLOBAL	MODALITES CARACTERISTIQUES	DES VARIABLES	IDEN	POIDS
				41.54	CLASSE 1 / 8		bb1b	211
14.79	0.000	63.50	98.10	64.17	married	marital status	MA02	326
14.11	0.000	62.96	96.68	63.78	male	sex of head of household	HH01	324
11.97	0.000	53.96	100.00	76.97	not stated nursing c	nursing costs for illness	IL04	391
11.97	0.000	53.96	100.00	76.97	labour performance n	labour performance	LA04	391
11.97	0.000	53.96	100.00	76.97	no illness	Disease	DI07	391
9.26	0.000	51.52	96.68	77.95	farming MA	most important activity	DM01	396
6.56	0.000	52.32	80.09	63.58	no orphans	keeping orphans?	OR02	323
6.56	0.000	51.80	81.99	65.75	no PLWHA	keeping person living with aids	HP02	334
5.96	0.000	82.98	18.48	9.25	fishing SA	second most important activity	DM03	47
3.93	0.000	57.50	32.70	23.62	medium-high AI	asset index	AS03	120
3.55	0.000	48.16	68.25	58.86	primary	highest level of education attained	ED02	299
3.22	0.001	51.08	45.02	36.61	no SA	second most important activity	DM15	186
2.83	0.002	57.97	18.96	13.58	more harvest	maize harvested	MA01	69
2.65	0.004	45.56	75.36	68.70	2 meals adults	adults meals	AD03	349
2.55	0.005	51.56	31.28	25.20	high AI	asset index	AS04	128
2.43	0.007	50.74	32.70	26.77	small LI	livestock ownership	LI01	136
-2.43	0.008	31.67	18.01	23.62	chilubi	district (qual)	DI01	120
-2.98	0.001	23.33	6.64	11.81	muchinshi	locality	LO08	60
-3.14	0.001	0.00	0.00	2.56	other, don't'know, w	Disease	DI06	13
-3.28	0.001	22.73	7.11	12.99	1 meal children	children meals	CH02	66
-3.35	0.000	33.33	36.49	45.47	missing area	maize area	M_04	231
-3.42	0.000	33.19	36.49	45.67	harvest not compared	maize harvested	MA04	232
-3.50	0.000	33.05	36.49	45.87	quantity sold not co	maize sold comparison	MA04	233
-3.60	0.000	0.00	0.00	3.15	Fever/malaria	Disease	DI02	16
-3.70	0.000	24.74	11.37	19.09	1 meal adults	adults meals	AD02	97
-3.75	0.000	0.00	0.00	3.35	Pneumonia, tb, skin	Disease	DI01	17
-4.11	0.000	4.00	0.47	4.92	post-secondary	highest level of education attained	ED04	25
-4.21	0.000	29.47	26.54	37.40	medium-high LI	livestock ownership	LI03	190
-4.53	0.000	0.00	0.00	4.53	household savings NC	nursing costs for illness	IL01	23
-4.65	0.000	0.00	0.00	4.72	old age, hypertensio	Disease	DI05	24
-4.95	0.000	14.71	4.74	13.39	no education	highest level of education attained	ED01	68
-5.31	0.000	0.00	0.00	5.91	formal employm MA	most important activity	DM13	30
-5.40	0.000	21.26	12.80	25.00	small AI	asset index	AS01	127
-5.73	0.000	0.00	0.00	6.69	unable to work	labour performance	LA01	34
-6.11	0.000	0.00	0.00	7.48	sale of crops NC	nursing costs for illness	IL02	38
-6.30	0.000	0.00	0.00	7.87	beer brewing SA	second most important activity	DM04	40
-6.49	0.000	0.00	0.00	8.27	fishing MA	most important activity	DM03	42
-6.56	0.000	21.84	18.01	34.25	yes PLWHA	keeping person living with aids	HP01	174
-6.56	0.000	22.70	19.91	36.42	yes orphans	keeping orphans?	OR01	185
-6.67	0.000	0.00	0.00	8.66	HIV/AIDS, STDs	Disease	DI04	44
-7.69	0.000	0.00	0.00	11.02	sale of livestock, h	nursing costs for illness	IL03	56
-7.93	0.000	0.00	0.00	11.61	divorced/separated	marital status	MA03	59
-8.16	0.000	2.70	0.95	14.57	farming SA	second most important activity	DM01	74
-9.13	0.000	0.00	0.00	14.76	below usual ability	labour performance	LA02	75
-11.77	0.000	0.00	0.00	22.44	widowed	marital status	MA04	114
-14.11	0.000	3.80	3.32	36.22	female	sex of head of household	HH02	184

CLASSE 2 / 8

V.TEST	PROBA	---- POURCENTAGES ---- CLA/MOD	MOD/CLA	GLOBAL	MODALITES CARACTERISTIQUES	DES VARIABLES	IDEN	POIDS
				9.06	CLASSE 2 / 8		bb2b	46
13.10	0.000	85.71	78.26	8.27	fishing MA	most important activity	DM03	42
12.15	0.000	55.41	89.13	14.57	farming SA	second most important activity	DM01	74
4.94	0.000	30.00	39.13	11.81	mumba	locality	LO05	60
4.21	0.000	12.88	91.30	64.17	married	marital status	MA02	326
3.96	0.000	19.17	50.00	23.62	chilubi	district (qual)	DI01	120
3.94	0.000	14.72	73.91	45.47	missing area	maize area	M_04	231
3.91	0.000	14.66	73.91	45.67	harvest not compared	maize harvested	MA04	232
3.88	0.000	14.59	73.91	45.87	quantity sold not co	maize sold comparison	MA04	233
3.48	0.000	12.35	86.96	63.78	male	sex of head of household	HH01	324
3.23	0.001	30.00	19.57	5.91	*Reponse manquante*	locality	861_	30
3.07	0.001	40.00	13.04	2.95	trading MA	most important activity	DM08	15
2.66	0.004	20.00	26.09	11.81	muchinshi	locality	LO08	60
2.65	0.004	15.83	41.30	23.62	mungwi	district (qual)	DI04	120
2.46	0.007	11.38	82.61	65.75	no PLWHA	keeping person living with aids	HP02	334
2.40	0.008	10.74	91.30	76.97	no illness	Disease	DI07	391
2.40	0.008	10.74	91.30	76.97	not stated nursing c	nursing costs for illness	IL04	391
2.40	0.008	10.74	91.30	76.97	labour performance n	labour performance	LA04	391
-2.36	0.009	0.00	0.00	9.25	fishing SA	second most important activity	DM03	47
-2.36	0.009	0.00	0.00	9.25	trading SA	second most important activity	DM08	47
-2.46	0.007	4.60	17.39	34.25	yes PLWHA	keeping person living with aids	HP01	174
-2.49	0.006	3.33	8.70	23.62	isoka	district (qual)	DI03	120
-2.71	0.003	4.59	19.57	38.58	less sold	maize sold comparison	MA03	196
-2.75	0.003	1.27	2.17	15.55	hhold chores SA	second most important activity	DM09	79
-2.80	0.003	0.00	0.00	11.61	lukulu	locality	LO02	59
-3.00	0.001	1.90	4.35	20.67	less area	maize area	M_03	105
-3.26	0.001	1.75	4.35	22.44	widowed	marital status	MA04	114
-3.48	0.000	3.26	13.04	36.22	female	sex of head of household	HH02	184
-3.75	0.000	0.00	0.00	17.52	finkuli	locality	LO01	89
-5.19	0.000	1.08	4.35	36.61	no SA	second most important activity	DM15	186
-5.31	0.000	0.00	0.00	29.13	mpika	district (qual)	DI02	148
-12.24	0.000	0.00	0.00	77.95	farming MA	most important activity	DM01	396

CLASSE 3 / 8

V.TEST	PROBA	---- POURCENTAGES ---- CLA/MOD	MOD/CLA	GLOBAL	MODALITES CARACTERISTIQUES	DES VARIABLES	IDEN	POIDS
				7.09	CLASSE 3 / 8		bb3b	36
13.10	0.000	96.67	80.56	5.91	formal employm MA	most important activity	DM13	30
9.52	0.000	80.00	55.56	4.92	post-secondary	highest level of education attained	ED04	25
7.04	0.000	38.46	55.56	10.24	more than 3 meals ad	adults meals	AD04	52
7.00	0.000	31.08	63.89	14.57	farming SA	second most important activity	DM01	74
5.65	0.000	23.86	58.33	17.32	more than 3 meals ch	children meals	CH04	88
4.70	0.000	10.74	97.22	64.17	married	marital status	MA02	326
4.23	0.000	23.33	38.89	11.81	muchinshi	locality	LO08	60
4.22	0.000	10.49	94.44	63.78	male	sex of head of household	HH01	324
2.70	0.003	13.33	44.44	23.62	chilubi	district (qual)	DI01	120
2.62	0.004	8.70	94.44	76.97	labour performance n	labour performance	LA04	391
2.62	0.004	8.70	94.44	76.97	no illness	Disease	DI07	391
2.62	0.004	8.70	94.44	76.97	not stated nursing c	nursing costs for illness	IL04	391
-2.33	0.010	0.00	0.00	11.61	divorced/separated	marital status	MA03	59
-2.60	0.005	0.00	0.00	13.39	no education	highest level of education attained	ED01	68
-3.06	0.001	3.85	30.56	56.30	2 meals children	children meals	CH03	286
-3.12	0.001	0.88	2.78	22.44	widowed	marital status	MA04	114
-4.22	0.000	1.09	5.56	36.22	female	sex of head of household	HH02	184
-4.37	0.000	3.44	33.33	68.70	2 meals adults	adults meals	AD03	349
-6.00	0.000	1.34	11.11	58.86	primary	highest level of education attained	ED02	299

-8.43	0.000	1.26	13.89	77.95	farming MA	most important activity	DM01	396

CLASSE 4 / 8

V.TEST	PROBA	POURCENTAGES CLA/MOD	POURCENTAGES MOD/CLA	POURCENTAGES GLOBAL	MODALITES CARACTERISTIQUES	DES VARIABLES	IDEN	POIDS
				8.66	CLASSE 4 / 8		bb4b	44
13.59	0.000	69.49	93.18	11.61	divorced/separated	marital status	MA03	59
8.58	0.000	22.83	95.45	36.22	female	sex of head of household	HH02	184
4.81	0.000	35.00	31.82	7.87	beer brewing SA	second most important activity	DM04	40
3.21	0.001	10.74	95.45	76.97	not stated nursing c	nursing costs for illness	IL04	391
3.21	0.001	10.74	95.45	76.97	no illness	Disease	DI07	391
3.21	0.001	10.74	95.45	76.97	labour performance n	labour performance	LA04	391
2.72	0.003	16.19	38.64	20.67	less area	maize area	M_03	105
2.42	0.008	10.82	84.09	67.32	15-49	age category	AG02	342
-2.35	0.009	2.27	4.55	17.32	same area	maize area	M_02	88
-2.48	0.007	1.35	2.27	14.57	farming SA	second most important activity	DM01	74
-2.52	0.006	1.33	2.27	14.76	below usual ability	labour performance	LA02	75
-2.58	0.005	1.30	2.27	15.16	65+	age category	AG04	77
-4.32	0.000	0.00	0.00	22.44	widowed	marital status	MA04	114
-8.58	0.000	0.62	4.55	63.78	male	sex of head of household	HH01	324
-9.07	0.000	0.31	2.27	64.17	married	marital status	MA02	326

CLASSE 5 / 8

V.TEST	PROBA	POURCENTAGES CLA/MOD	POURCENTAGES MOD/CLA	POURCENTAGES GLOBAL	MODALITES CARACTERISTIQUES	DES VARIABLES	IDEN	POIDS
				12.40	CLASSE 5 / 8		bb5b	63
12.19	0.000	49.12	88.89	22.44	widowed	marital status	MA04	114
10.51	0.000	32.61	95.24	36.22	female	sex of head of household	HH02	184
8.52	0.000	29.19	85.71	36.42	yes orphans	keeping orphans?	OR01	185
7.47	0.000	45.59	49.21	13.39	no education	highest level of education attained	ED01	68
4.92	0.000	15.86	98.41	76.97	not stated nursing c	nursing costs for illness	IL04	391
4.92	0.000	15.86	98.41	76.97	labour performance n	labour performance	LA04	391
4.92	0.000	15.86	98.41	76.97	no illness	Disease	DI07	391
3.35	0.000	32.50	20.63	7.87	beer brewing SA	second most important activity	DM04	40
3.34	0.000	25.32	31.75	15.55	hhold chores SA	second most important activity	DM09	79
3.20	0.001	21.26	42.86	25.00	small AI	asset index	AS01	127
2.95	0.002	14.65	92.06	77.95	farming MA	most important activity	DM01	396
-2.56	0.005	0.00	0.00	7.48	sale of crops NC	nursing costs for illness	IL02	38
-2.68	0.004	1.79	1.59	11.02	sale of livestock, h	nursing costs for illness	IL03	56
-2.81	0.002	1.69	1.59	11.61	divorced/separated	marital status	MA03	59
-2.84	0.002	0.00	0.00	8.66	HIV/AIDS, STDs	Disease	DI04	44
-3.06	0.001	9.06	49.21	67.32	15-49	age category	AG02	342
-3.14	0.001	8.36	39.68	58.86	primary	highest level of education attained	ED02	299
-3.42	0.000	1.35	1.59	14.57	farming SA	second most important activity	DM01	74
-4.12	0.000	0.00	0.00	14.76	below usual ability	labour performance	LA02	75
-8.52	0.000	2.79	14.29	63.58	no orphans	keeping orphans?	OR02	323
-9.90	0.000	1.53	7.94	64.17	married	marital status	MA02	326
-10.51	0.000	0.93	4.76	63.78	male	sex of head of household	HH01	324

CLASSE 6 / 8

V.TEST	PROBA	POURCENTAGES CLA/MOD	POURCENTAGES MOD/CLA	POURCENTAGES GLOBAL	MODALITES CARACTERISTIQUES	DES VARIABLES	IDEN	POIDS
				5.31	CLASSE 6 / 8		bb6b	27
11.65	0.000	91.67	81.48	4.72	old age, hypertensio	Disease	DI05	24
7.40	0.000	27.27	77.78	15.16	65+	age category	AG04	77
7.06	0.000	32.14	66.67	11.02	sale of livestock, h	nursing costs for illness	IL03	56
7.03	0.000	44.12	55.56	6.69	unable to work	labour performance	LA01	34
3.67	0.000	13.16	55.56	22.44	widowed	marital status	MA04	114
3.66	0.000	16.00	44.44	14.76	below usual ability	labour performance	LA02	75
2.66	0.004	9.14	62.96	36.61	no SA	second most important activity	DM15	186
2.57	0.005	8.95	62.96	37.40	medium-high LI	livestock ownership	LI03	190
2.56	0.005	13.24	33.33	13.39	no education	highest level of education attained	ED01	68
-2.57	0.005	0.00	0.00	17.32	more than 3 meals ch	children meals	CH04	88
-2.75	0.003	3.07	37.04	64.17	married	marital status	MA02	326
-6.05	0.000	0.88	11.11	67.32	15-49	age category	AG02	342
-8.84	0.000	0.00	0.00	76.97	labour performance n	labour performance	LA04	391
-8.84	0.000	0.00	0.00	76.97	not stated nursing c	nursing costs for illness	IL04	391
-8.84	0.000	0.00	0.00	76.97	no illness	Disease	DI07	391

CLASSE 7 / 8

V.TEST	PROBA	POURCENTAGES CLA/MOD	POURCENTAGES MOD/CLA	POURCENTAGES GLOBAL	MODALITES CARACTERISTIQUES	DES VARIABLES	IDEN	POIDS
				4.72	CLASSE 7 / 8		bb7b	24
8.93	0.000	69.57	66.67	4.53	household savings NC	nursing costs for illness	IL01	23
8.89	0.000	87.50	58.33	3.15	Fever/malaria	Disease	DI02	16
6.33	0.000	13.22	95.83	34.25	yes PLWHA	keeping person living with aids	HP01	174
6.25	0.000	22.67	70.83	14.76	below usual ability	labour performance	LA02	75
2.99	0.001	17.50	29.17	7.87	beer brewing SA	second most important activity	DM04	40
2.80	0.003	15.91	29.17	8.66	HIV/AIDS, STDs	Disease	DI04	44
2.52	0.006	15.79	25.00	7.48	sale of crops NC	nursing costs for illness	IL02	38
2.41	0.008	9.65	45.83	22.44	widowed	marital status	MA04	114
-6.33	0.000	0.30	4.17	65.75	no PLWHA	keeping person living with aids	HP02	334
-8.27	0.000	0.00	0.00	76.97	not stated nursing c	nursing costs for illness	IL04	391
-8.27	0.000	0.00	0.00	76.97	no illness	Disease	DI07	391
-8.27	0.000	0.00	0.00	76.97	labour performance n	labour performance	LA04	391

CLASSE 8 / 8

V.TEST	PROBA	POURCENTAGES CLA/MOD	POURCENTAGES MOD/CLA	POURCENTAGES GLOBAL	MODALITES CARACTERISTIQUES	DES VARIABLES	IDEN	POIDS
				11.22	CLASSE 8 / 8		bb8b	57
11.43	0.000	79.55	61.40	8.66	HIV/AIDS, STDs	Disease	DI04	44
10.63	0.000	54.67	71.93	14.76	below usual ability	labour performance	LA02	75
9.49	0.000	58.93	57.89	11.02	sale of livestock, h	nursing costs for illness	IL03	56
8.78	0.000	28.74	87.72	34.25	yes PLWHA	keeping person living with aids	HP01	174
8.06	0.000	63.16	42.11	7.48	sale of crops NC	nursing costs for illness	IL02	38
6.79	0.000	82.35	24.56	3.35	Pneumonia, tb, skin	Disease	DI01	17
5.40	0.000	21.74	70.18	36.22	female	sex of head of household	HH02	184
4.90	0.000	25.44	50.88	22.44	widowed	marital status	MA04	114
4.14	0.000	38.24	22.81	6.69	unable to work	labour performance	LA01	34
3.60	0.000	53.85	12.28	2.56	other, don't'know, w	Disease	DI06	13
3.51	0.000	27.12	28.07	11.61	lukulu	locality	LO02	59
2.91	0.002	20.62	35.09	19.09	1 meal adults	adults meals	AD02	97
2.41	0.008	22.03	22.81	11.61	divorced/separated	marital status	MA03	59
2.38	0.009	16.89	43.86	29.13	mpika	district (qual)	DI02	148
-2.55	0.005	0.00	0.00	8.27	fishing MA	most important activity	DM03	42
-2.60	0.005	1.67	1.75	11.81	kampumbu	locality	LO04	60
-2.69	0.004	2.60	3.51	15.16	65+	age category	AG04	77
-5.40	0.000	5.25	29.82	63.78	male	sex of head of household	HH01	324
-6.05	0.000	4.60	26.32	64.17	married	marital status	MA02	326
-8.78	0.000	2.10	12.28	65.75	no PLWHA	keeping person living with aids	HP02	334
-13.67	0.000	0.00	0.00	76.97	no illness	Disease	DI07	391
-13.67	0.000	0.00	0.00	76.97	not stated nursing c	nursing costs for illness	IL04	391
-13.67	0.000	0.00	0.00	76.97	labour performance n	labour performance	LA04	391