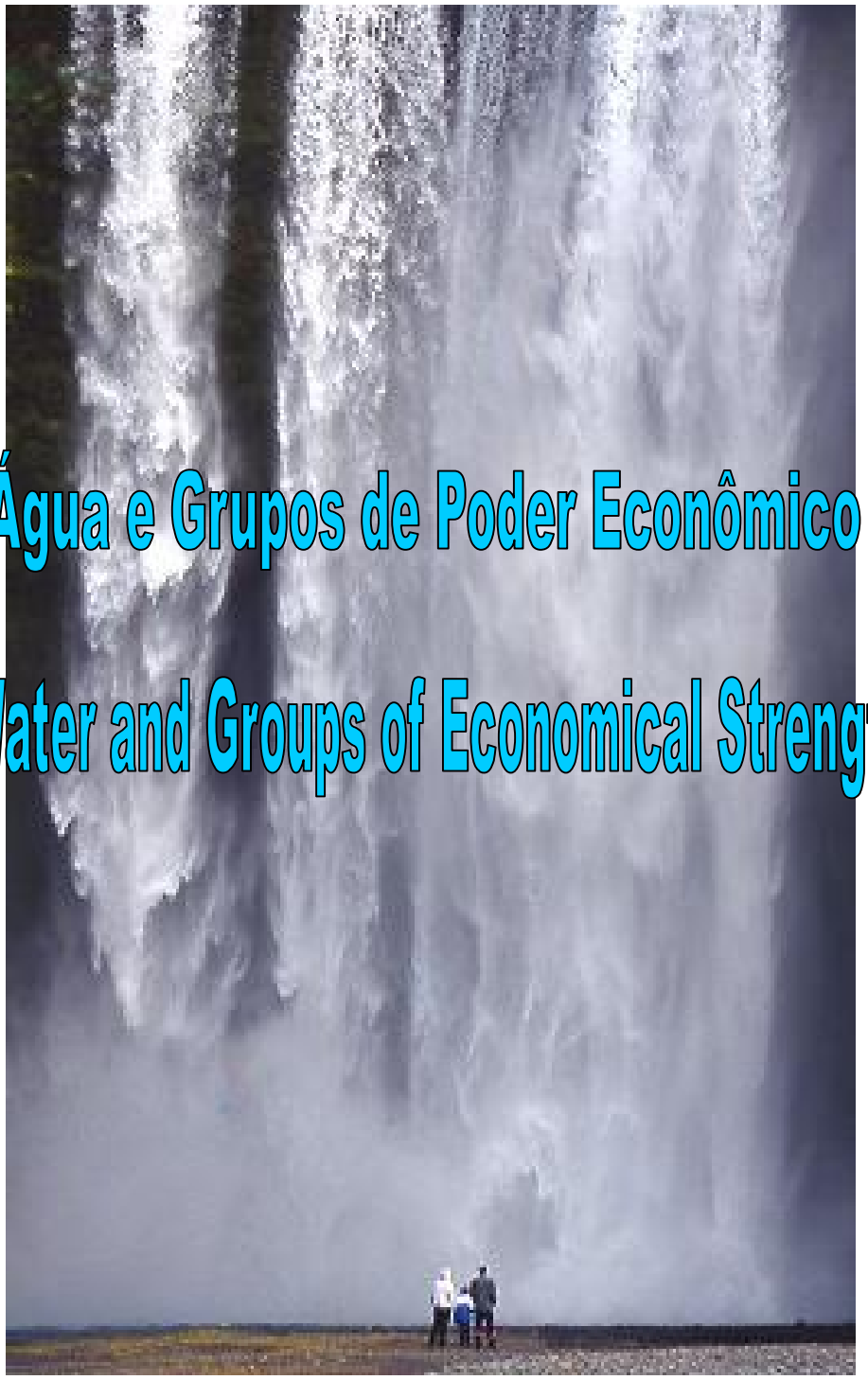
Klaus Röder -Consultant (CIM) at the National Statistical Institute (INE) Mozambique, stationed at Beira, Capital of the Province of Sofala
Assessor/ Consultant (GTZ-CIM) do Instituto Nacional da Estatística (INE)- Região Central
Cellular/Mobile:00258-82-5374670
Email: Kroeder@gmx.net



Uso de Água e Grupos de Poder Econômico em Maputo

Use of Water and Groups of Economical Strength in Maputo

Beira, Monday, 04 June 2007

Table of Contents

Foreword

This document is part of the results of consulting contract between SAL “Consultoria em Desenvolvimento Social Lda” and the author.

One of the objectives of this work was to analyse the behaviour of families in the Metropolitan Area of Maputo concerning the consumption of Water. This behaviour, it is felt to be heavily dominated by economical aspects. The first task was to define groups of economical status, usually called wealth groups. The data derived from the analysis of the survey should not be confounded with poverty figures derived from Household Surveys [6]. Whereas the later poverty figures are based on consumption figure, the index of economical strength is a proxy index based on household assets available. For the examples the version SPSS 10.0.7 was used.

The Classification of Groups of Economical Strength

How to classify groups of economical strength or status if no direct data of income or expenditure are available. This paper follows the approach of Fimer and Pritchett[7] very precisely. This will enable the reader to follow up he methodology in more detail in that document although the data set and purpose of it is quite different. The task of classifying wealth is not an easy one. Many lines of statistical text has been written how to measure wealth/poverty. Looking at a set of available data a very simplistic albeit reasonable approach is to look at the assets the household owns to define its wealth group. All analysis takes into consideration only the data of the selected sample. Like in every inductive statistics it is assumed that the sample is randomly selected and represents the behaviour of the universe. Fimer and Pritchett propose a principal component approach to define wealth groups. This approach uses the factor analysis procedure to calculate a wealth index (called asset index in our example). Descriptive statistics are calculated for this asset index to calculate mean and standard deviation and the quintiles dividing the sample into 5 almost equally sized groups named as below:

		Asset Groups					Total
		Very Few Assets	Few Assets	Neither Few nor Many Assets	Many Assets	Very Many Assets	
Cidade	Maputo	73	89	82	92	85	421
	Matola	46	35	34	36	28	179
Total		119	124	116	128	113	600

The groups are further combined to the Lower 40% comprising group “Very Few Assets and Few Assets” , the Middle 40% comprising “Neither Few nor Many Assets and Many Assets” and the Upper 20% owning “Many Assets” This Asset Index (AI) is not better than any other poverty index but it is easy to calculate, robust and a good indicator of long –run wealth.
Basic knowledge of factor analysis with principal components is required to understand the process of generating the AI and the Asset Groups. In order to explain how these Asset groups are calculated a short explanation is necessary although more comprehensive explanation is left to specialized text-books [2][17].

Some basic principles of Factor Analysis

In social and economical analysis the researcher is often confronted with complex terms: “Intelligence”, “Entrepreneurship”. “Economical wealth and status” ca also be regarded as such. It would be possible to analyse special variables to determine these complex terms The Factor Analysis takes into consideration many variables and tries to find out, if there is a hidden “background” variable common to the observed variables and not directly visible. These “background” variables are called factors. It is the aim of the factor analysis to reduce the complexity of information to isolate few common factors. Once running the factor analysis, the result resembles an equation of regression:

$$ExpVarF = a_1 * Fac1 + a_2 * Fac2 * + .. + a_n * Facn + \mathrm{Re}\, s$$

where *ExpVarF* is the Explained Variable which contrary to the Regression approach does not appear in the data set. The “Variables” *Fac1..n* corresponding to the explaining variables of the Regression are called the (common) Factors or components,(if derived by the method of principal components, these are called components) or explaining sets of variables to be named by the analyst. The coefficients a1..n corresponding to the Regression coefficients are called the factor loadings and explain in numerical terms how much a Factor contributes to the explanation of the explained Variable. The variable Res is the residual containing the unexplained rest. Let us repeat: Factors and ExpFacF are unknown entities, which as a result of the Factor analysis and the judgement of the analyst become named entities. Explaining the steps to calculate the Factors in the below mentioned example might show more clearly the meaning of these denominations.

The Asset Index: Factor Analysis extracting only one factor

The Asset Index is a simplified calculation of a factor analysis, considering ExpVarF the long- run material wealth of the respondents or in short *EconomicalStrength*. The only factor to be taken into consideration is the first factor or the other factors have been deliberately excluded from the analysis is the first one and this is by choice the assets being available in the household or let us name it simply *Assets*. So the Formula in our example reads:

$$EconomicalStrength = a \ * Assets + Res$$

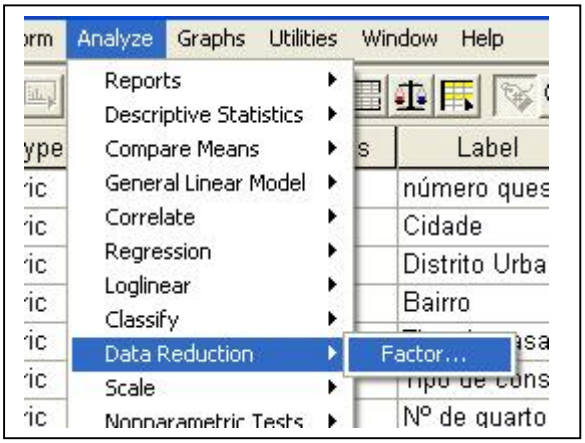
So like in Regression this formula is a vector calculation, so that this calculations really means

$$EconomicalStrength = \begin{pmatrix} a_1 \\ a_2 \\ .. \\ a_n \end{pmatrix} * (Asset_1, Asset_2, ..., Asset_n) + Res$$

Formula 1

..and so the value of *EconomicalStrength* of each household is calculated by multiplying the factor loadings of the n assets with the corresponding values of each asset in the household.

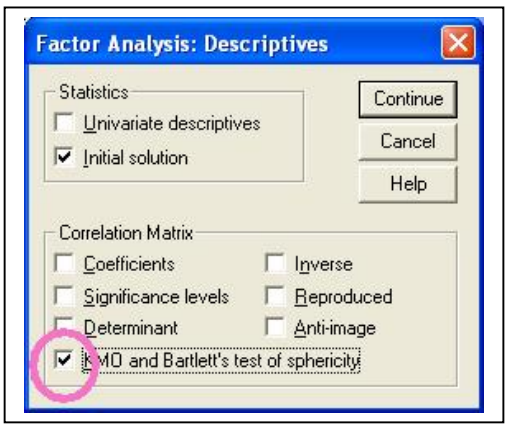
How is it done in SPSS:



(p6) and the asset variables P108 until P129

After opening the data set, call the command <Analyze/Data Reduction/Factor…)

The next dialog box lets you choose the variables to be included in the Factor analysis. Let us choose the Number of Rooms used

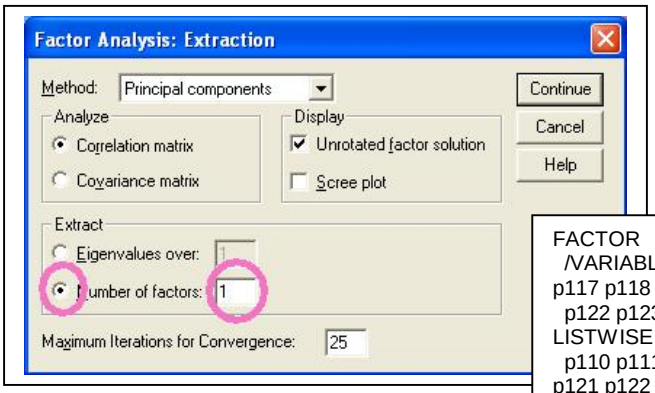


except the variable P116 (“Tem computador”). These are 21 variables. This choice will be explained later.

Some choices have to be made in the different subsequent dialog boxes, for example, choose “KMO and Bartlett´s test of sphericity” in the “Descriptives”-box.

The KMO or Kaiser-Meyer-Olkon measure will calculate an indicator of the quality of the factor model. The results will be discussed later in detail.

In the following “Extraction” box you have to choose the number of factor to be extracted as 1. This is not the normal choice because generally more factors than one are tentatively extracted but in our example and following the explanation of Fimer and Pritchett[7] the one and only factor will suffice.



No other alteration in the dialog boxes is necessary and the pasted SPSS command then look like this

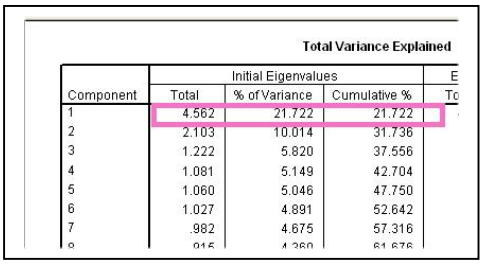
```
FACTOR
  /VARIABLES p6 p109 p110 p111 p112 p113 p114 p115
p117 p118 p119 p120 p121
  p122 p123 p124 p125 p126 p127 p128 p129  /MISSING
LISTWISE /ANALYSIS p6 p109
  p110 p111 p112 p113 p114 p115 p117 p118 p119 p120
p121 p122 p123 p124 p125
  p126 p127 p128 p129
  /PRINT INITIAL KMO EXTRACTION
  /CRITERIA FACTORS(1) ITERATE(25)
  /EXTRACTION PC
  /ROTATION NOROTATE
  /METHOD=CORRELATION .
```

No rotation is necessary (and possible) because only one factor only is extracted.
As a result we first observe the following box

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.845
Bartlett's Test of Sphericity	Approx. Chi-Square	2408.018
	df	210
	Sig.	.000

showing a KMO measure of 0.845. Any KMO measure above 0.8 indicates a good choice of variables for the factor analysis anything of a measure below 0.5 would recommend a different choice of variables. The KMO combines the indiactors of sample adequacy as can be asked for in the “Descriptives” box, clicking the Anti-Image option. There are analysts who reduce the number of variables by dropping any variable with a diagonal value of less than (e.g. 0.66) but for the sake of simplicity and with much more theoretical discussion all the 21 chosen variables will be maintained



Total Variance Explained

Component	Initial Eigenvalues			E
	Total	% of Variance	Cumulative %	To
1	4.562	21.722	21.722	
2	2.103	10.014	31.736	
3	1.222	5.820	37.556	
4	1.081	5.149	42.704	
5	1.060	5.046	47.750	
6	1.027	4.891	52.642	
7	.982	4.675	57.316	
8	.915	4.360	61.676	

		N° de quartos	Tem ligação energia eléctrica	Tem fogão	Tem aspirador de pó	Tem lava-louça	Tem ferro engomar
Anti-image Covariance	N° de quartos	.773	-8.860E-02	2.201E-03	6.584E-02	-3.892E-02	7.556E-03
	Tem ligação energia eléctrica	-8.860E-02	.529	-3.279E-03	5.250E-02	1.711E-03	-7.462E-02
	Tem fogão	2.201E-03	-3.279E-03	.673	-7.772E-03	-.126	-.127
	Tem aspirador de pó	6.584E-02	5.250E-02	-7.772E-03	.779	-5.091E-02	1.901E-02
	Tem lava-louça	-3.892E-02	1.711E-03	-.126	-5.091E-02	.831	-5.990E-02
	Tem ferro engomar	7.556E-03	-7.462E-02	-.127	1.901E-02	-5.990E-02	.553
	Tem máquina lavar louça	7.052E-04	-1.868E-02	1.313E-02	5.440E-02	-5.493E-02	7.014E-02
	Tem video	-4.083E-02	-2.019E-02	-6.685E-03	-4.611E-02	-2.334E-02	-8.552E-02
	Tem micro ondas	-3.887E-03	2.104E-02	-.122	-.151	-9.926E-02	2.839E-02

Having accepted the model we shall remain looking at he results of the "Communalities" and "Eigenvalues", denominations crucial to the understanding of the Factor analysis. The following table indicates how much of the total standard deviation of all variables will be explained by the component/ factor. This means that the first factor "explains" 4.318 23% of the total deviation (all variables are automatically transformed to z-scores: normalized to mean 0 and standard deviation 1 before). In a more stringent and fully executed Factor analysis, the five factors/components with values higher than 1 would probably be accepted in the model, but the the question of explaining the "factors" would be more difficult. This might be easier to understand if one looks at the final result of the Factor analysis, the Component matrix. These values for the only chosen component indicate in which respect the chosen variable (being all part of factor1) contributes to the calculation of the explained variable *EconomicalStrength.* and these are the values to be multiplied by the values of the respective household to calculate the value of the Asset Index (asindx1).or the variable *EconomicalStrength* like in Formula 2, which is the more explicit form of Formula 1, where f_1 is the first scoring factor (f_N is last) as defined in the Component Matrix, a_{ji} *is* the j[th] household value for the i[th] asset and $\overline{a_i}$ *and* s_i are the mean and standard deviation of the i[th] asset for all households

Component Matrix[a]

	Component 1
N° de quartos	.481
Tem ligação energia eléctrica	.668
Tem fogão	.631
Tem aspirador de pó	.324
Tem lava-louça	.435
Tem ferro engomar	.704
Tem máquina lavar louça	.152
Tem video	.553
Tem micro ondas	.444
Tem vontoinha	.633
Tem ar condicionado	.339
Tem congelador	.604
Tem máquina de costura	.234
Tem telefone fixo	.278
Tem celular	.391
Tem rádio	.303
Tem televisão	.592
Tem geleira	.515
Tem bicileta	.209
Tem motocicleta	.013
Tem carro	.506

Extraction Method: Principal Component Analysis.

a. 1 components extracted.

There have been some attempts without the room numbers and with the additional variable P116 ("tem computador"). That variable P116 yields a negative factor loading which means this asset contributes negatively to the *EconomicalWealth* or in other words having a computer makes

$$ExpVarF = f_1 * (a_{j1} - \overline{a_1})/(s_1) * ..... * f_N * (a_{jN} - \overline{a_N})/(s_N)$$

Formula 2

```
compute assindx1= 0.481*(P6-2.292)/(0.921)+0.66*(P109-0.6)/(0.49)+0.646*(P110-
0.188)/(0.391)+0.338*(P111-0.015)/(0.122)+0.438*(P112-0.227)/(0.419)+0.711*(P113-
0.368)/(0.483)+0.157*(P114-0.015)/(0.122)+0.56*(P115-0.248)/(0.432)+0.454*(P117-
0.037)/(0.188)+0.639*(P118-0.308)/(0.462)+0.345*(P119-0.023)/(0.151)+0.592*(P120-
0.372)/(0.484)+0.24*(P121-0.087)/(0.282)+0.298*(P122-0.053)/(0.225)+0.392*(P123-
0.808)/(0.394)+0.322*(P124-0.787)/(0.41)+0.589*(P125-0.707)/(0.456)+0.521*(P126-
0.197)/(0.398)+0.219*(P127-0.093)/(0.291)+0.045*(P128-0.017)/(0.128)+0.511*(P129-
0.087)/(0.282).
```

Explicit Formula 2 as in SPSS Syntax file

the Household poorer (?). This does not make much sense. The reason for this behaviour lies in the fact that a full Factor analysis would yield different factors like “Basic Assets” and “Sophisticated Assets”. The computers will probably contribute positively to this second component together with assts like “maquina a lavar louça” and “micro onda”.For the sake of simplicity this issue has not been followed up further

The Asset Groups are calculated simply by asking for the quintiles of the variable *EconomicalStrength.*
Group values are attributed according to the quintiles and the new variable Assetgr (Asset Groups or Groups of Economical Strength are calculated accordingly)

Statistics

Asset Index w.NoRooms

N	Valid	588
	Missing	12
Mean		9.652E-02
Std. Deviation		4.6071
Percentiles	20	-4.2942
	40	-1.5807
	60	.7015
	80	3.6435

```
FREQUENCIES
  VARIABLES=assindx1
/FORMAT=NOTABLE
  /NTILES= 5
  /STATISTICS=STDDEV MEAN
  /ORDER=  ANALYSIS .
```

```
*Group definition 1.
IF (assindx1 <= -4.294) assetgr1 = 1 .
IF (assindx1 > -4.294 & assindx1 <= -1.581) assetgr1 = 2 .
IF (assindx1 > -1.581& assindx1 <= 0.701) assetgr1 = 3 .
IF (assindx1 > 0.701& assindx1 <= 3.644) assetgr1 = 4 .
IF (assindx1 > 3.644) assetgr1 = 5.
```

The Classification of Population Groups by Income and Expenditure

The grouping of the population according to Income and Expenditure is the more classical procedure. However this method is less straightforward than the asset based classification described above.
The sources of information are the IAF data of the National Household Survey 2002/2003 [18]. These data can be analysed on Province level, which means the data for the City of Maputo can be evaluated. There is no information available on “Bairro” Level.
The information however have to be considered with caution for several reasons. For various reasons Income information have the reputation to be rather unreliable. This leads to the classical exploitation of expenditure data to determine income groups.

The following data have been compiled from the IAF data set from Maputo City. The data are extrapolated, the net number of households is 623, the extrapolated number is 148115 so the sample stands roughly for 150 thousand households in Maputo City.

Income by Household.

N	Valid	148115
	Missing	745
Mean		4187
Median		1571

This result shows an average monthly income of 4187 Meticais. This differs slightly from the publication [18] as it does not include transfers. Transfers have been omitted because of their rather arbitrary and spotty nature. Moreover the data have been cleaned of the 10% outliers at the top level because these deteriorate heavily the results.

The five quintiles are defined by the following Income Limits

Income Group	Income Limits MTN	Mean of Income MTN	Std. Error
1		490	1.43
2	> 700 AND <= 1245	1154	1.19
3	> 1245 AND <= 2250	1995	2.12
4	> 2250 AND <= 5000	3905	5.75
5	> 5000	15726	69.79

That means that the households of the lowest income group have an income of less than 700 Meticais per month, the highest income group disposes of an income of more than 5000 Meticais

The per Capita Income figures show the following

Mean		796
Median		500
	Income Group	Income Limits
	1	<= 211
	2	> 211 AND <= 400
	3	> 400 AND <= 716
	4	> 716 AND <= 1220
	5	> 1220

That means the average per capita income in Maputo is about 800 Meticais and the group limits are as indicated.

As it comes to expenditure, the data show that expenditure is somewhat higher than the income. How is this possible? One explanation would be that, that the respondents underestimated their income, which is a plausible motive.

The expenditure mean and quintile groups by Households

Mean	4634
Median	1952

Expenditure Group	Expenditure Limits	Mean of Expenditure MTN	Std. Error
1	<= 865	343	2.78
2	> 865<= 1488	960	0.90
3	> 1488<= 2594	1683	1.66
4	> 2594<= 5533	3300	4.57
5	> 5533	14899	86.65

This means that a Maputo Households has an average monthly expenditure of roughly 4600 Meticais, the poorest Households have no more than 870 Meticais to spend, the least poor have more than 5500 Meticais.

Remember that the 10% top Outliers have been removed so the data differ slightly from [18], in fact they are lower than the data displayed there. On the individual level the same table reads like

Mean		764
Median		353
	Expenditure Group	Expenditure Limits
	1	<= 140
	2	> 140<= 266
	3	> 266<= 492
	4	> 492<= 1140
	5	> 1140

So an average individual expenditure of 764 Meticais has been calculated for Maputo City and an individual belonging to the lowest income group has less than 140 Meticais to spend per month. These figures include income and expenditure in money and kind.

As a final consideration: In Maputo the expenditure for water supply and services of habitation, like garbage removal (“Abastecimento de água e serviços de habitação”) represents 1.61 Percent of the total expenditure in Maputo [19]. This is expenditure group “044” and it has been calculated for December 2004, based again on the data collected in the National Household Survey 2002/2003 [18].

References

[1] SPSS 10.0 Manuals - SPSS Inc. Headquarters, 233 S. Wacker Drive, 11th floor Chicago, Illinois 60606
[2] SPSS 10.0 Regression Models ISBN 0130179043
[3] SPSS 10.0 for Windows Student Version ISBN 0130280402
[4] Discovering Statistics Using SPSS for Windows : Advanced Techniques for Beginners (Introducing Statistical Methods series) by Andy Field (Paperback)
[5] How to Lie With Statistics by Darrell Huff, Irving Geis
[6] Pobreza E Bem-Estar Em Moçambique: Segunda Avaliação Nacional - Março de 2004 Direção Nacional do Plano e Orçamento, Ministério de Plano e Finanças Gabinete de Estudos, Ministério de Plano e Finanças Instituto Internacional de Pesquisa em Políticas Alimentares (IFPRI) Universidade de Purdue
[7] Deon Filmer / Lant Pritchett 2004 Estimating Wealth Effects without Expenditure Data -- or Tears: An Application to Educational Enrollments in States of India
[8] Deaton, A., and S. Zaidi. 1999. Guidelines for constructing consumption aggregates for welfare analysis. Research Program in Deve1opment Studies Working Paper No. 192. Princeton, N.J.: Princeton University.
[9] J.-L-Dubois, D.Blaizeau - Connaître les conditions de vie de ménages dans les pays en développement –Tome 3: Analyser les résultats- - Ministère de la Coopération et du développement- ISBN 2-11-ß84855-3
[10]Foster, J., J. Greer, and E. Thorbecke. 1984. A class of decomposable poverty measures. Econometrica
[11]52 (3): 761-765.
[12]Ravallion, M. 1994. Poverty comparisons. Chur, Switzerland: HarwoodAcademic Publishers.
[13]-' 1998. Poverty lines in theory and practice. Living Standards Measurement Study Working Paper No. 133. Washington, D.C.: World Bank.
[14]-' 2001. Growth, inequality and poverty: Looking beyond averages. Policy Research Working Paper No. 2558. Washington, D.C.: World Bank.
[15]Tarp, E, C. Arndt, H. T. Jensen, S. Robinson, and R. Heltberg. 2002a. Facing the development challenge in Mozambique: An economywide perspective. Research Report 126. Washington, D.C.: International Food Policy Research Institute.
[16]Tarp, F., K. R. Simler, C. Matusse, R. Heltberg, and G. Dava. 2002b.. Economic Development and Cultural Change 51 (1): 77-108.
[17] Gerhard Brosius 1989 SPSS- Advanced Statistics and Tables McGraw-Hill Book Company Inc.
[18] Inquérito aos Agregados Familiares sobre Orçamento Familiar – Relatório Final
[19] ÍNDICE DE PREÇOS NO CONSUMIDOR - CIDADE DE MAPUTO (BASE: DEZEMBRO 2004=100) Instituto Nacional de Estatística