Tropical Agricultural Research Vol. 12: 287-296 (2000)

Estimation of Food and Nutrition Factor in Human Development

A.P.G.S. De Silva, R.O. Thattil[1] and S. Samita[1]

Postgraduate Institute of Agriculture
University of Peradeniya
Peradeniya, Sri Lanka

ABSTRACT. *This study explains how weighted principal components technique is applied to variables which can be structured in some meaningful way into sets so that the principal components of the respective sets extract maximum variance. The method is basically a process of forming factors out of variables which are adjusted to their variances and eigen values within the structured sets of variables. The study illustrates how the technique is used to obtain two underlying factors and a composite index (CI) out of the variables which have been used to compile the Human Development Report by the United Nations statistical office. A comparison between the human development index (HDI) and the CI has also been done. The study reveals that though Sri Lanka is ranked high in HDI with substantially higher Gross Domestic Product (GDP) value and education achievement indicators, its food production and nutritional status of the people are not in par with the other human development indicators.*

INTRODUCTION

The United Nations Development Programme (UNDP) started an exercise to compare the level of human development between member countries since 1990. Four variables namely life expectancy (LE), adult literacy rate (ALR) and gross enrolment ratio (GER), and adjusted GDP (AGDP) are used to compute the HDI. The variables cover three aspects of human development namely health, education and income. Since the process applied by the UNDP takes into consideration only simple averages of respective relative indices, an alternative technique which takes into account mean, variation and the correlation structure of the variables will be of immense value, to put HDI in its proper perspective. The HDI based ranking of nations has a lot to be desired.

This study aims at extending the scope of the HDI by adding two more variables to represent the food and nutrition level of the country, which is an important aspect of human development. Food production per capita index and per capita calorie supply were found to be representative of the food and nutrition aspect. The multivariate technique, namely the factor analysis, which was used for analysis, enabled the estimation of a food and nutrition factor in addition to the human development factor, as already represented by the HDI.

1 Department of Crop Science, Faculty of Agriculture, University of Peradeniya, Peradeniya, Sri Lanka.

METHODOLOGY

More than sixty variables were selected from the list of variables which was available in the Human Development Report (United Nations Development Programme, 1998). A preliminary scrutiny exercise was carried out to short list the number of variables down to a manageable size, by taking the correlation, high dependency and causal factors among the variables into consideration. Originally there were about 62 variables of 173 member countries structured in more than 10 sets. The selection of countries had to be restricted to developing countries since the defined variables differed between developed and developing countries. In selecting variables, since the data availability was a prime concern in using the covariance option, the variables for which values were not available for all the countries were omitted. At the same time countries with a lot of missing variables too were omitted. Finally, a data matrix of 19 variables and 69 countries was selected for analysis. The structuring of variables was done as described earlier considering the relevance and correlation of the variables within a set. Number of sets finally structured was four with ten variables in all sets.

A basic factor analysis was performed in order to study the movements of the contributions from each variable (communalities) to different factors formed. Using this we can identify the variables with similar movements so that it can be used to group the variables in addition to the use of the correlation pattern.

This study used the weighted principal component technique to arrive at a composite index using factor analysis (Morrison, 1976). Usually, principal components are constructed by using the correlation coefficients of the original variables. It is equivalent to the use of standardized variables. In other words the weights used in respect of all variables are equal. The covariance option in the PRINCOMP procedure of SAS (1987) can be used to construct weighted principal components using user specified weights. Forming a new variable by dividing the original variable with its standard deviation (σ_{ij}), we can apply equal weights to the principal components thus constructed with the use of covariance option.

$$x_{ij}^* = x_{ij} / \sigma_{ij} \text{ ; for all i and j}$$

(same result is obtained when $x_{ij}^* = \{x_{ij} - E(x_{ij})\}/\sigma_{ij}$ is used)

where x_{ij}^* is the new variable and x_{ij} is the original variable (subscript j indicates the set while i indicates variables within each set).

Now we can introduce a weight (w_j) for each j^{th} set of variables using, $w_j = 1/\lambda_j$, where λ_j is the eigen value of the first Principal Component of the set j. The new variable will thus become,

$$x_{ij}' = x_{ij}^* / \sqrt{\lambda_j} \qquad \text{(Garcia } \textit{and Puetra}\text{, 1997)}$$

When we apply factor analysis to the new variables x_{ij}', the weights are applied to the original x_{ij} accordingly. Dividing by σ_{ij} we standardize the variables while dividing by λ_j we strike a balance between the sets. This will maintain the equal importance between the sets as well.

The weight for each set is $w_j = k / \lambda_j$ (k times $1/\lambda_j$), where k is the number of variables in the j^{th} set. All these parameters can be changed and tested for different models.

When the data matrix $\underline{X}$ is expressed in terms of derived factors ($\underline{f}$):

$\underline{X} = \underline{\Lambda} . \underline{f} + \underline{e}$ (Anderson, 1984)

Λ representing the matrix of factor loadings (λ_{ij}) in respect of each X_j variable such that,

$X_j = \Lambda_j . f + e_j$, e_j being the unexplained portion.

The communality of the X_j is derived by,

$V(X_j) = \sum^{m}_{k=1} \lambda^2_{jk} + \Psi_j$, where Ψ_j is the specific variance of X_j.

In this study we attempt to maximize the communalities of variables using structured variable approach as described in the following section.

RESULTS AND DISCUSSION

Final analysis was based on 10 variables relating to 69 countries in the developing world. The following variables were selected for structuring into new sets.

X1	-	Life expectancy at birth - LE (years) 1995
X2	-	Infant mortality rate - IMR (per 1000 live births) 1996
X3	-	Adult literacy rate (%) - ALR 1995
X4	-	Gross enrolment ratio for all levels (%) - GER 1996
X5	-	Real GDP per capita (PPP$) 1995
X6	-	Daily per capita supply of calories 1995
X7	-	Food production per capita index (1980 = 100) 1996
X8	-	Crude birth rate - CBR 1995
X9	-	Women's share of labour force 1995
X10	-	Percentage of labour force in agriculture 1990

The communality changes as the number factors are increased from 1 to 4. The results indicate that there is a convergence in the communalities of all the variables as the number of factors are increased (Table 1). The variables X1, X2, X8 and X10 have moved approximately the same way. Therefore, we can remove the variables X2, X8 and X10 without changing the structure of the model, keeping X1 which has the highest communality within first set in the first two factors. Then, if we want a single factor obtained, the best choice of variables are X1, X4 and X5. We can also note that UNDP has opted to choose X1, X3, X4 and X5 in its HDI. Variable X3 can also contribute to factor one. If we want to use two factor analysis, then the improved communality of variables X6

Table 1. Change in communality when number of factors is increased.

Variable	Factor 1	Factor 1 & 2	Factor 1, 2 & 3	Factor 1, 2, 3 & 4
X1	.79	.81	.81	.82
X2	.76	.81	.81	.81
X8	.75	.76	.82	.84
X3	.60	.81	.90	.90
X4	.65	.78	.78	.84
X5	.67	.67	.67	.81
X6	.55	.80	.82	.90
X7	.25	.69	.93	.99
X9	.33	.42	.84	.95
X10	.77	.78	.85	.86

and X7 can be judged as a reflection of importance. Variable X9 can only be included if we use three factors. Model improves substantially if we use four factors. More factors we use, it will be more difficult to describe the factors in an independent manner. The other interesting thing to note is the unaltered communality of X5 between the three factors (0.67).

By judging the improvements in the communality values, it can also be determined the nature and the structure of the factors. Thus, the third factor would reflect X7 and X9 which can be described as women's share in the labour force in food production.

In this study we used the two factor approach. The final analysis is therefore based on the six variables contributing to the two factors (excluding X2, X8, X9 and X10). The X5 has a constant correlation to all factors which moves closely with the variable X1, and therefore, can be grouped together in further analysis. The new structure of variables thus belong to two sets namely;

a) human development

and

b) food production and nutrition

In restructuring the variables, the principal components of each set was constructed so as to get only a single principal component with eigen value greater than one

(1.0). In moving the variables, care was taken so that a particular set will have meaningful relationships between the variables of the set. The value of the first eigen value was maximized (Table 2).

Table 2. The final groupings of the variables and corresponding eigen values.

Eigen value of first PC (λ_{1j})	% variance explained	Group name/set	Variables in set	Weight k / λ_j
2.99287	75	Human development	X1, X3, X4, X5	1.33650
1.53145	77	Food production and nutrition	X6, X7	1.30595

The human development set contains the variables X1, X3 , X4 and X5; while the food production and nutrition set contains X6 and X7. The variables were then pooled back to perform factor analysis using square root of the eigen values as the weights to re-scale the variables. The results of the weighted factor analysis is given in the Table 3.

Table 3. Results of the weighted factor analysis.

Eigen value	Percentage	Cumulative %
1.52	57.6	57.6
0.54	20.5	78.1
0.30	11.1	89.3
0.12	4.6	93.9

It is noted that only 58% of variability is explained by the first factor. This indicates that there are other important factors, which can explain substantial amount of variability among the selected variables. When the two factor model is considered a total of 78% is accounted for by the model. The Table 4 gives the correlation between the factors and variables, together with corresponding communalities.

Table 4. Correlation between factors and variables.

Variable	F_1	F_2	Communality
X1	0.81	0.34	0.77
X3	0.68	0.54	0.76
X4	0.71	0.52	0.77
X5	0.80	0.24	0.70
X6	0.84	-0.22	0.76
X7	0.67	-0.64	0.86

However, the correlation structure does not warrant a clear interpretation due to contributions from variables by X3, X4 and X7 into both factors. Therefore, a rotation was done in order to come up with a better correlation pattern between factors and variables. The results of the varimax rotation is given in the Table 5.

Table 5. Correlation between factors (rotated) and variables.

Variable	F_1	F_2
X1	0.82	0.31
X3	0.87	0.08
X4	0.87	0.12
X5	0.75	0.38
X6	0.46	0.74
X7	0.05	0.93

Communality values suggest that all variables were adequately explained in the two factor model with exceptionally high contribution from X7 which is food production. Thus it can be said that food production aspect plays a major role in explaining the variation between the countries according to the selected model.

Equation of standardized variables for factor 1 is:

$$F_1 = 0.27\,X1 + 0.34\,X3 + 0.33\,X4 + 0.23\,X5 + 0.08X6 - 0.32X7$$

This factor explains the 'human development' of a country with higher correlations and approximately equal positive coefficients in respect of variables X1, X3, X4 and X5. Variables X6 and X7 both have either very low correlations or smaller coefficients, thus minimising the effect of the two variables in the factor.

The correlation between the first factor and the HDI of UNDP is estimated at 0.90. This shows the similarities between the two indices. Also, due to the application of a proper multivariate technique to arrive at the model, the index represented by the first factor can be considered superior. It can also be mentioned that some of the differences are due to the use of direct GDP values in the factors as opposed to Atkinson adjustment used on GDP in HDI. While the effect of the adjustment is very little for the developing countries as compared to the developed countries, the developing countries with high GDP values are better represented here.

Equation of standardized variables for factor 2 is:

$$F_2 = -0.03\,X1 - 0.14\,X3 - 0.12\,X4 = 0.01\,X5 + 0.44\,X6 + 0.76\,X7$$

Except for variables X6 and X7, rest of the variables have correlations lesser than 0.4 and coefficients lesser than 0.2. It can be observed that the segregation of variables between the two factors, which is a sign of good factoring.

Calorie supply and food production can be treated as indicators of current nutritional status of people in the country. Thus, the second factor clearly describes a separate dimension of development in the above form. We also noted low correlations between the second factor with LE and GDP. The factor can be named as 'Food supply and nutritional status' factor. This factor has a correlation of 0.35 only with HDI. This shows that the HDI is not a good indicator of food and nutrition status of the people. The combined index can be formed using the two factors as follows:

CI = F1 + F2 (since the two factors explain almost equal variations of 1.05 and 1.00 respectively).

The final rankings can be based on the CI. The ranking show that the following countries are in the front: UAE, Korea, Lebanon, China, Chile, Uruguay, Argentina, Mauritius, Malaysia and Turkey.

Change for the better based on the CI rather than HDI are the countries such as UAE, Indonesia, China, Egypt, Vietnam, Morocco and Pakistan due to their improved food and nutrition level (also for certain countries such as UAE due to use of direct GDP values

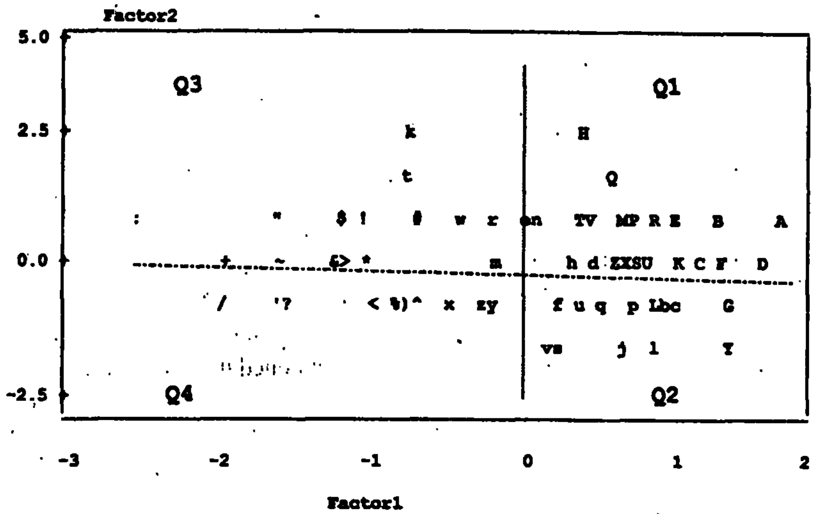

Fig. 1. Plot of factor 1 and factor 2. #
[Note: # Q1....Q4 represent different clusters]

symbol	country *	symbol	country *	symbol	country *
A	Korea (2)	Y	Cuba (41)	w	Myanmar (33)
B	Chile (5)	Z	Peru (31)	x	Cameroon (49)
C	Costa Rica (14)	a	Jordan (25)	y	Lesotho (53)
D	Argentina (7)	b	Dominican (35)	z	Kenya (56)
E	Uruguay (6)	c	Sri Lanka (36)	!	Pakistan (43)
F	Fiji (11)	d	Paraguay (26)	#	India (40)
G	Panama (29)	e	Indonesia (21)	$	Nigeria (39)
H	UAE (1)	f	Botswana (44)	%	Togo (60)
J	Mexico (13)	g	Philippines (34)	^	Zambia (61)
K	Colombia (20)	h	Guyana (30)	&	Cote d'ivorie (48)
L	Thailand (32)	j	Mongolia (54)	*	Mauritania (57)
M	Malaysia (9)	k	China (4)	(	Tanzania (64)
N	Mauritius (8)	l	Namibia (47)	)	Yemen (63)
O	Brazil (16)	m	Guatemala (42)	<	Madagascar (65)
P	Belize (17)	n	Egypt (18)	>	Sudan (59)
Q	Lebanon (3)	o	El Salvador (37)	~	Senegal (52)
R	Turkey (10)	p	Swaziland (38)	`	Malawi (62)
S	Ecuador (27)	q	Honduras (45)	"	Guinea-Bissau (50)
T	Iran (15)	r	Vietnam (28)	'	Mozambique (69)
U	Syria (23)	s	Solomon (55)	?	Guinea (67)
V	Algeria (19)	t	Morocco (22)	+	Mali (66)
W	Tunisia (12)	u	Nicaragua (46)	:	Burkina Faso (58)
X	Jamaica (24)	v	Zimbabwe (51)	/	Niger (68)

* countries are listed vertically down according to HDI ranking.
* within parenthesis are the ranking based on the CI

as against Atkinson values). Change for the worse based on the CI than HDI are the countries such as Fiji, Panama, Cuba, Dominican Rep., Sri Lanka, Philippines and Mongolia due to their poor food and nutrition level.

The countries could be represented on a two factor plane. The countries with balanced development will be mostly explained by the positive quarters of the two factors while where there are imbalances in development, the second and third quarters will explain it. The fourth quarter will explain the under-development (Fig. 1)

The above plot indicates the variation of development of countries according to the above criteria. Extent of development is explained by the first factor. Countries with high human resource development (HRD) are mostly explained by the first factor. Food and nutrition status is explained by the second factor. Countries such as UAE (H), Lebanon (Q), China (k) and Morocco (t) have low HRD values compared to their food status values thus indicating an imbalance nature of development. Countries such as Cuba (Y), Namibia (I), Panama (G), Sri Lanka (c) and Mongolia (j) are countries with higher HRD values and lower food and nutritional status values. They can be treated as less improved in food and nutritional status but with high status of health and educational achievements.

The ranking of Sri Lanka according to the study is 36 as compared to that of 28 in HDI, out of the 69 selected countries in the developing countries category.

CONCLUSIONS

The weighted principal component technique could be effectively applied to the variables as given in the UNDP HDI report, in order to construct factor based composite indices. The factors thus constructed better represent the variations in the human development among the countries. The structuring of variables as explained in the article can be further studied and developed in order to arrive at efficient models in weighted principal component analysis using desirable weights for each variable concerned. It can also be concluded that the HDI itself has logical explanations in the present formulation. The omission of variables such as Food production and Calories supply reduced the efficiency of the HDI in adequately explaining the variations of human development between the countries. It can be concluded that, though Sri Lanka is high in the HDI rank with substantially higher GDP value and education achievement indices, its food production and nutritional status of people are not in par with other human development indicators.

Thus, the estimated factors explain a country's standing in terms of LE, education, food production, nutrition and GDP, while the HDI only explains the standing in terms of LE, education and GDP.

REFERENCES

Anderson, T.W. (1984). An introduction to multivariate statistical analysis. 2nd edition, Willy & Sons Inc., New York.

Garcia Del, V.I.T. and Puetra, G.C. (1997). Human development: weighted principal components analysis applied to the data of 1997 report. Joint IASS/IAOS Conference Proceedings, Mexico.

Morrison, D.F. (1976). Multivariate statistical methods, 2nd edition, McGraw-Hill Book Co., New York.

SAS. (1987). Statistical analysis system, SAS/STAT guide, SAS Institute Inc., Cary, North Carolina.

United Nations Development Programme. (1998). Human Development Report, New York.