Determinants of a Better Living
- A Cross-Country Analysis -

Abu Jafar Mohammad Sufian

The Population Crisis Committee, Draper Fund rated 130 countries on the International Human Suffering Index created on the basis of 10 socioeconomic, and development variables. The countries' have then been classified into three groups: group 1 - high living standard countries, group 2 - medium living standard countries, and group 3 - low living standard countries, depending on their index values. The canonical discriminant analysis technique has been employed in this paper to identify those variables which contribute most in discriminating among the three groups. The analysis shows that the first canonical variable accounts for 92 percent of the total discriminable variance existing in the discriminating variables. In addition, the analysis has correctly classified 120 out of 130 countries lending support to the adequacy of the discriminant analysis. Also, an examination of both the standardized coefficients and the total structure coefficients shows that the infant mortality is the most influential variable in the first function. Among the variables considered in this study, only the inflation variable failed to contribute substantially in discriminating the groups. Policy implications are discussed.

Author: Department of Urban and Regional Planning, College of Architecture and Planning, King Faisal University, Saudi Arabia.

Ⅰ. Introduction

The Population Crisis Committee (PCC), Draper Fund - a committee concerned with solutions of world population problems - created an index - International Human Suffering Index (Hoffman and Hoffman Public Relations, 1987) on the basis of the following 10 measures of human welfare: gross national product per capita in US$ (abbreviated as GNP), average annual rate of inflation (INFLAT), infant mortality rate (INFMORT), daily per capita calorie supply as percent of requirement (CALORIE), percent of population with access to clean drinking water (CLWATER), energy consumption per capita in Gigajoules (ENERGY), adult literacy rate (LITERACY), average annual growth of labour force (LGROWTH), average annual growth of urban population (UGROWTH), and personal freedom. Each of these 10 variables was ranked from 0 to 10 - the most distressful being 10.

The PCC rated the living conditions of 130 countries - countries where the relevant data were available. A country with high infant mortality rate, low gross national product per capita, poor supplies of clean drinking water etc., scored high on the index - close to 100. The 10 measures used in the index were judged to be the best among those available. A number of other measures deemed relevant to the index were considered for inclusion in the index, but all had some weakness, or were duplicated measures already chosen, or did not cover enough countries. The countries rated in the index were divided into four groups - minimal human suffering, moderate human suffering, high human suffering, and extreme human suffering - according to their index values. Switzerland scored the least on the index, implying that human suffering is the least in this country, while Mozambique scored the highest. These ratings of living condition show that the majority of the world's people have to endure lives of human misery, measured in terms of the index.

A question that follows relates to whether the different groups of countries differ in respect of all the 10 variables that were used in creating the index. It may so happen that the groups do differ only on some measures but not on others. In that case, the areas of greater attention will concern those variables on which the groups differ. It is, therefore, important to determine which, if any, of those variables are useful in predicting the ultimate fate of the countries with low level of living.

Thus, the aim of this paper is to identify those variables which are most powerful, that is, which contribute most, in discriminating among the different groups of countries having different living standards.

II. Data and Methods

A. Data

The same data used by the PCC have analysed in this paper. The following is the list of sources of data cited in the Index sheet: The World Bank (The World Development Report, 1986), United Nations International Children's Fund (State of the World's Children, 1984 and 1986, 1986 and 1987), Food and Agriculture Organization (The State of Food in Agriculture, 1984), U.S. Department of Agriculture, Embassy of Iceland, Washington D.C., Global Water, Washington D.C., The United Nations (Energy Statistics Yearbook, 1984), Freedom House (Freedom at Issue, January~February, 1987). Some of the estimates used by the PCC and subsequently used in this analysis were made by the PlanEcon, Inc. and varied U.S. Government Sources. The Population Crisis Committee itself made some estimates which were based on interviews with government officials and development experts from private organizations. Population size and the rate of natural increase data were the mid 1986 estimates from 1986 world population data sheet, Population Reference Bureau.

B. Variables

The variable 'personal freedom' has been dropped from the present analysis altogether. This had to be done since this variable is neither a ratio nor even an interval measure - a prerequisite for the applicability of the technique called dscriminant analysis which has been employed in this paper. The new index (that is, the index formed by excluding the 'personal freedom' variable) ranging from 0 to 90 has now been employed for yielding new groups for analysis. Countries having ties in the index values have been treated as if they have different indices. The groups are as follows:

Group	Index Value	Number of Countries in Group	Description
1	0~37	40	Countries with high living standard
2	37~64	50	Countries with medium living standard
3	64~90	40	Countries with low living standard

Another variable measuring annual population increase in percent (PGROWTH) which was not used in creating the index but whose values appeared on the index sheet, has been used in the present analysis. Thus, in all, 10 variables, in addition to the clssification variable which is the index itself, have been used in this analysis. The variables and their measurers have been discussed in detail in the index sheet.

C. Analytical Technique

The analytical technique used in this research is the canonical discriminant analysis. The technique is appropriate when there are two or more groups which can be presumed to differ on several variables, measurable at least at the interval level, and the objective is to answer the question of whether it is possible to discriminate among the groups on the basis of the variables under consideration, an if so, how well do they discriminate, and which variables are the most powerful discriminators. It is also applied for the purpose of classification of a case with unknown group membership but with known values on the discriminating variables (Klecka, 1980; Nie et al., 1975).

The assumptions on which the technique rests are that no variable is a linear combination of other discriminating variables and that the variables within groups have multivariate normal distributions with equal covariance matrices, i.e.,

$$f(\underset{\sim}{X},\ \mu_i, \Sigma) = \frac{1}{(2\pi)^{P/2}\,|\Sigma|^{1/2}}\, e^{-\frac{1}{2}(\underset{\sim}{x}-\mu_i)'\Sigma^{-1}(\underset{\sim}{x}-\mu_i)}$$

where $\underset{\sim}{x}$ is the vector of variables

$\underset{\sim}{x}' = (x_1, x_2 \ldots x_p)$

μ_i is the mean vector of the ith group $\mu'_i = (\mu_{i1}, \mu_{i2}, \ldots \mu_{ip})$

and Σ is the common covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \ldots\ldots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \ldots\ldots & \sigma_{2p} \\ \vdots & \vdots & & \vdots \\ \sigma_{p1} & \sigma_{p2} & \ldots\ldots & \sigma_{pp} \end{bmatrix}$$

where σ_{ii} is the variance of the ith variable while σ_{ij} is the covariance between the ith and the jth variables. The technique requires the derivation of canonical discriminant functions which are functions linear in the discriminating variables having the form

$$S_{km} = \alpha_0 + \alpha_1 X_{1km} + \alpha_2 X_{2km} + \ldots + \alpha_p X_{pkm} \quad \ldots\ldots(1)$$

where S_{km} is the score on the canonical discriminant function for case m in group K, X's are discriminating variables, and α's are the desired coefficients and are called unstandardized coefficients since the discriminating variables, the X's, have not been measured in standard form. If the X's are measured in standard form, then the associated coefficients are the standardized coefficients.

The α's are derived for the first function such that all the group means on the function are maximally discriminated. It may distinguish between some of the groups but not others in which case the second canonical discriminant function is derived such that it best separates the groups, with the proviso that it is orthogonal or uncorrelated with the first. Further canonical discriminant functions are derived such that they also provide maximal separation between the groups whilst being uncorrelated with the previous discriminant functions (Bennet and Bowers, 1976).

The derivation of the coefficients involves solving the simultaneous equations

$$(B-\lambda W)\underset{\sim}{\beta}=\underset{\sim}{0}$$

where B and W are the between and within - groups sums of squares and cross product matrices, $\underline{\beta}$ is a vector of p coefficients, and λ is the eigenvalue. The coefficients in $\underline{\beta}$ are then adjusted to get α's by using

$$\alpha_i=\beta_i\sqrt{n.-g}$$

and

$$\alpha_o=-\sum_{i=1}^{p}\alpha_i X_{i..}$$

where n. = total number of cases over all the groups

g = number of groups

$X_{i..}$ = mean of the ith variable over all the cases.

The standardized coefficients are then derived by using the following transformation

$$\gamma_i=\alpha_i\sqrt{\frac{w_{ii}}{n.-g}}$$

where w_{ii} is the within sum of squares of variable i.

Findings

Table 1 shows the mean values of the discriminating variables for the three groups.

As expected, group 1 (high living standard countries) has the highest

Table 1. Mean Values of the Discriminating Variables for the Three Groups of Countries - High Living Standard, Medium Living Standard and Low Living Standard

Variable	Group 1 (High Living Standard)	Group 2 (Medium Living Standard)	Group 3 (Low Living Standard)
GNP	8,088.40	1,957.96	374.98
INFLAT	12.39	18.49	19.56
INFMORT	14.75	60.70	121.88
CALORIE	125.00	110.08	90.19
CLWATER	93.53	63.42	28.45
ENERGY	136.68	29.20	4.16
LITERACY	94.44	74.00	50.25
PGROWTH	0.68	2.54	2.79
LFORCE	0.87	2.76	2.69
UGROWTH	1.85	4.12	5.15

Note: All computer work in this paper have been done by using the Statistical Analysis System Package.

average GNP (US$8,088) while group 3 (low living standard countries) has the lowest average (US$375); and group 2 falls at intermediate position (US$1,958). In general, the groups appear to have different values on each variable. It might, therefore, be anticipated that the variables will do well in discriminating the groups.

Also Wilk's lambda defined as

$$\Lambda = \prod_{i=r+1}^{q} \frac{1}{1+\lambda_i}$$

where r denotes the number of functions already derived, q denotes the maximum number of functions, and λ_i is the eigenvalue associated with the ith discriminant function, is a multivariate measure of group differences over several discriminating variables. The eigenvalues are given in table 2 below.

Before any functions have been derived, i.e., when r=0, we have Λ=0.056222. Because lambda is an inverse measure, the above value of lambda implies that the 10 variables used are extremely effective in differentiating among the groups.

Table 3 reports the standardized discriminant coefficients for the two canonical discriminant functions.

Before interpreting the coefficients let us first judge the importance of the two derived canonical discriminant functions which are as follows:

$$S_{1km} = 0.2468\ Z_1 + 0.0208\ Z_2 - 0.8641\ Z_3 + 0.4203\ Z_4 + 0.5806\ Z_5 + 0.5350\ Z_6 + 0.3200\ Z_7 - 0.4218\ Z_8 - 0.3469\ Z_9 + 0.0150\ Z_{10} \quad \text{................} \quad (2)$$

$$S_{2km} = 0.2023\ Z_1 - 0.0225\ Z_2 - 0.7485\ Z_3 - 0.5500\ Z_4 + 0.4820\ Z_5 + 0.7510\ Z_6 - 0.3336\ Z_7 - 1.0217\ Z_8 - 0.1701\ Z_9 - 0.0844\ Z_{10} \quad \text{...............} \quad (3)$$

where the Z's are X's expressed in standard forms.

Table 2 provides necessary information to make this judgement. The sum

Table 2. Eigenvalues and Canonical Correlations

Discriminant Function i	Eigenvalue λ_i	Relative Percentage	Canonical Correlation R_i
1	9.0940	92.27	0.945419
2	0.7621	7.73	0.632939

of the eigenvalues is a measure of the total variance existing in the discriminating variables. The table shows that the first function has accounted for 92 percent of the total discriminable variance, implying that this first function acts highly efficiently in separating the three groups. The second function contains only about eight percent of the total discriminating power in this system of equations.

A further aid in judging the substantive utility of a discriminant function is by examining its associated canonical correlation coefficient symbolised as R, which is related to the eigenvalue as

$$R_i = \sqrt{\frac{\lambda_i}{1+\lambda_i}}$$

This coefficient measures how closely the function and the 'group variable' are related. This provides another

Table 3. Standardized Canonical Coefficients

	First Function	Second Function
GNP	0.2468	0.2023
INFLAT	0.0208	-0.0225
INFMORT	-0.8641	0.7485
CALORIE	0.4203	-0.5500
CLWATER	0.5806	-0.4820
ENERGY	0.5350	0.7510
LITERACY	0.3200	-0.3336
PGROWTH	-0.4218	-1.0217
LFORCE	-0.3469	-0.1701
UGROWTH	0.0150	-0.0844

Table 4. Number of Observations and Percents Classified into Different Groups

Groups	1	2	3	Total
1	37 92.50	3 7.50	0 0.00	40 100.00
2	0 0.00	45 90.00	5 10.00	50 100.00
3	0 0.00	2 5.00	38 95.00	40 100.00
Total	37	50	43	130
Percent	28.46	38.46	33.08	100.00
Priors	0.3333	0.3333	0.3333	

measure to judge the function's ability to discriminate among the groups. Table 2 shows that the first discriminant function is strongly related (R=.95) to the groups. The second function is also correlated, though not strongly, with the group variable.

Another way of judging the adequacy of the derived discriminant functions, as well as of assessing how effective the discriminating variables are, is to classify the cases which are used in deriving the functions. Table 4 shows the classification of the countries into different groups. Clearly, the proportion of correctly classified cases is quite large (120 out of 130 i.e., 92.3 percent), implying the success of the discriminant analysis.

Further evidence about the group differences can be derived from the group centroids in the space of the discriminant functions (Table 5) and their corresponding plot (Figure 1).

Table 5. Mean Discriminant Function Scores

Variable	Group 1 (High Living Standard)	Group 2 (Medium Living Standard)	Group 3 (Low Living Standard)
First Discriminant Function	4.0725	-0.4878	-3.4627
Second Discriminant Function	0.5341	-1.0822	0.8187

The diagram suggests that all three groups are widely separated on the first discriminant function. On the second function, although groups 1 and 3 are hardly separated, group 2 is distinguished, although moderately.

To interpret these discriminant functions on the basis of the original variable set, we need to look at the standardized coefficients given in table 3. These coefficients are of great analytic value in and of themselves.

When the sign is ignored, each coefficient represents the relative contribution of its associated variable to that function. The sign denotes whether the contribution of the variable is positive or negative. Thus we notice for function 1, the most influential variable is the infant mortality (coefficient is -0.8641). The negative sign indicates that as the infant mortality decreases,

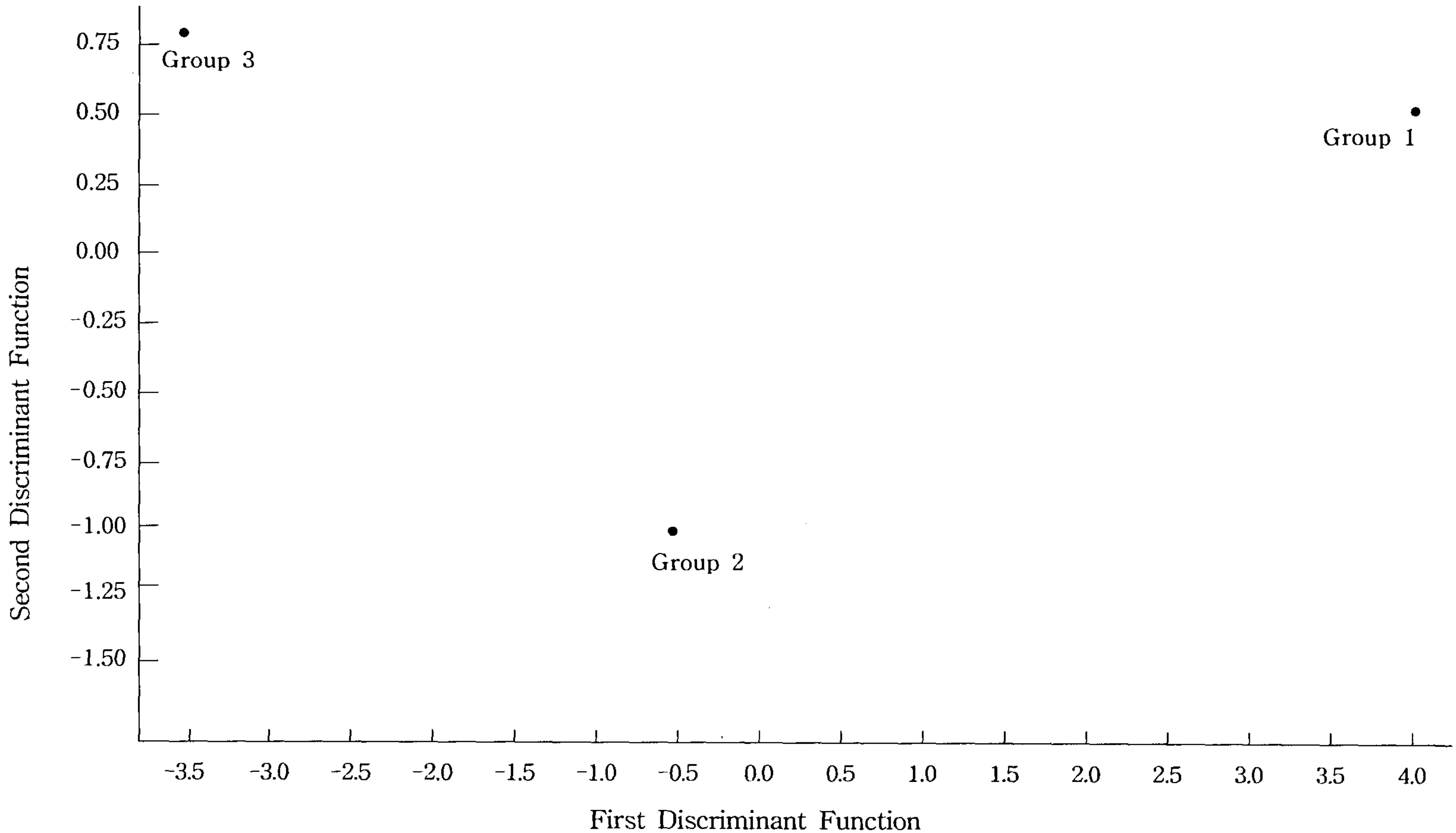

Figure1. Graphical Display of Discriminants Function Means

the value of the discriminant function increases and the likelihood that the country will belong to group 1 increases.

This finding is consistent with results from previous researches. Infant mortality is one of the most revealing measures of how well a society is meeting the needs of its people (Newland, 1981: 5). In comparisons among societies, it is considered an indicator of deprivation. It is often a cause of concern about poverty (Food Research and Action Group, 1984), and about spending on social welfare program (Miller, 1985). It has also been proposed as one of the three equally weighted indicators as an adequate measure of minimum human needs (Morris, 1979).

Table 1 shows that the mean infant mortality rates of the countries in group 3 (low living standard countries) and of the countries in group 2 (medium living standard countries) are respectively eight and four times higher than that of the countries in group 1 (high living standard countries). It is, therefore, likely that infant mortality will be one of the most dominant variables in discriminating among the groups. The result is, thus, in the expected direction.

Infant mortality is followed by population having access to clean drinking water (0.5806), energy consumption (0.5350), population growth rate (-0.4218), calorie supply (0.4203), labour force (-0.3469), literacy rate (0.3200), and gross national product (0.2468) in that order. The positive sign of the access to clean drinking water, for example, indicates that as the percentage of those having access to clean drinking water increases, so does the value of the discriminant function. Inflation rate (0.0208) and urban growth (0.0150) are of very minor importance compared to other variables.

Thus, as the infant mortality, population growth rate, and growth of labour force decrease, or population having access to clean drinking water, energy consumption, calorie supply, literacy rate, and gross national product increase, the country is likely to have a high standard of living. However, as all variables except two, seem to contribute significantly to the discriminant function, it is difficult to ascribe any meaningful interpretation to it.

In the case of second function, the population growth rate contributes the most (-1.0217). The negative sign implies that as the growth rate increases the value of the discriminant function decreases, and therefore, the likelihood that the country will belong to group 2 increases. Energy consumption and infant mortality make almost similar contributions (0.7510 and 0.7485 respectively).

The next contributors are calorie supply (-0.5500), population having access to clean drinking water (-0.4820), literacy rate (-0.3336), gross national product (0.2023), and growth of labour force (-0.1701). The other two variables - inflation (-0.0225) and urban growth (-0.0844) remain unimportant as in the first function. As in the case of function 1, since most variables contribute significantly, we cannot ascribe any meaningful interpretation to this function also.

Table 6. Total Canonical Structure

	First Function	Second Function
GNP	0.7634	0.2473
INFLAT	-0.1982	-0.0826
INFMORT	-0.9147	0.2920
CALORIE	0.8156	-0.2609
CLWATER	0.8877	-0.2280
ENERGY	0.8499	0.2982
LITERACY	0.8462	-0.2175
PGROWTH	-0.8358	-0.3891
LFORCE	-0.7314	-0.5268
UGROWTH	-0.8053	-0.1157

Table 6 reports the total structure coefficients that tell us how closely a variable and a function are related. A large coefficient implies that the function is carrying nearly the same information as the variable. If two variables are highly correlated, they share their contribution to the discriminant score, and as a result, the standardized coefficients may be smaller than when only one of the variables is used, or may be larger but with opposite signs so that the balance of the contributions is retained. The structure coefficients are unaffected by relationships with other variables.

We notice that all but the inflation variable have high structure coefficients on the first function. Gross national product and urban growth variables had low standardized coefficients (0.2468 and 0.0150 respectively) but their structure coefficients are quite large (0.7634 and -0.8053 respectively). This may be due to a high correlation between gross national product and each of infant mortality (-0.5974), population growth (-0.5769), labour force (-0.5250), and urban growth (-0.5058); and between urban growth and each of calorie supply (-0.5854), supply of clean drinking water (-0.6706), energy consumption (-0.5939), and literacy rate (-0.7590). What happens here is that the variables highly correlated with gross national product or urban growth are making large negative contributions to the discriminant score while gross national product or urban growth are making small positive contribution so that the net effects represent their true effects upon the score. Total structure coefficients on the second function can be interpreted in a similar way. Thus, an examination of the total structure coefficients shows that all the discriminating variables considered in this study, except the inflation, do play significant roles in discriminating the three groups.

Ⅲ. Summary and Conclusions

The Population Crisis Committee developed an index - Human Suffering Index - based on data mainly from the World Bank and the United Nations on the following ten variables : gross national product per capita in US dollar, average annual rate of inflation, average annual growth of labour force, average annual growth of urban population, infant mortality rate, daily per capita calorie supply as percent of requirement, percent of population with access to clean drinking water, energy consumption per capita, adult literacy rate, and personal freedom. Then they rated 130 countries - countries where the relevant data were available - on the basis of this index. For the purpose of present analysis, the variable - personal freedom - has been dropped out of the index because it is not an interval measure - which is a requirement for the analytical technique used in this analysis. Based on these slightly changed index values, the countries have been stratified into three groups : group 1 - countries with high living standard (40 countries), group 2 - countries with medium living standard (50 countries), and group 3 - countries with low living standard (40 countries).

The canonical discriminant analysis technique has been used to identify the variables which contribute most in discriminating among the groups, and also to find functions to reduce the number of dimensions of the space where the data cases or their centroids could be located.

In addition to the nine variables mentioned above, the annual population growth has also been used as a discriminating variable.

An examination of the mean of the discriminating variables for the three groups, and the Wilk's lambda shows evidence that the groups do differ significantly among themselves. Since there are three groups, two canonical discriminant functions are possible. The eigenvalues associated with the two discriminant functions show that the first canonical variable accounts for 92 percent of the total discriminable variance existing in the discriminating variables. An examination of the canonical correlation also lends support to the overriding importance of the first function. Finally, the adequacy of the discriminant analysis has been judged through classification of the same 130 countries with known group affiliation. The percentage of correctly classified cases is over 92, implying that a considerable amount of success has been achieved in discrimination. The plot of the centroids in the reduced space of discriminant functions also shows that the groups are widely separated, particularly on the first discriminant function.

An examination of the standardized coefficients shows that the infant mortality is the most influential variable in the first function, followed by percent of population having access to clean drinking water, and energy consumption. Except inflation rate and urban growth, all other variables do contribute substantially to the discriminant score.

Population growth rate is the dominant variable in the second function. This is followed by energy consumption and infant mortality - the variables which contribute almost equally to the discriminant score. As in the first function, inflation rate, and urban growth do not contribute substantially to the discriminant score.

An examination of the total structure coefficients indicates that only the inflation variable has failed to contribute substantially in discriminating the groups. This also shows that the variables - gross national product, and urban growth - do also play important roles in discriminating the groups. The standardized coefficients did not manifest these influences, probably because of these two variables' high correlation with those variables which contribute to the discriminant scores substantially but in opposite directions. The net effects of these variables remained low. Structure coefficients, being unaffected by these relationships, have shown the importance of these variables.

One limitation of the study is that the criterion for differentiating the groups was not very firm. For example, in group 1 there are 40 countries. this number could be 39 or 41 without disturbing any firm criterion. But except at the borderlines, there are strong grounds for broadly differentiating the groups on the basis of the index values. Another limitation is that the classification criterion for testing the adequacy of the discriminant function is based on probabilities of group membership, the calculation of which is appropriate when the discriminating variables follow a multivariate normal distribution. The multivariate normality assumption has not been examined here. However, as Klecka (1980: 62) has observed, a high percentage of correct classification (which is very high in our case, 92.3 percent) shows that the violation of assumptions, if any, has not been very harmful. Only marginal improvements might be achieved through efforts to improve the data or through use of alternative formulas. Also several authors have shown that the discriminant analysis technique is very robust and the assumptions need not be strongly adhered to (Nie, et al., 1975, Lachenbruch, 1975)

The most significant finding of this analysis is that the infant mortality is the most important variable in discriminating the three groups. The situation demands that the socio-economic correlates of infant mortality at the national level be identified, and the factors that appear to be the most important determinants of infant mortality be paid proper attention. Since this subgroup of the population - infants - is very vulnerable to the environmental conditions, attempts sought to reduce infant mortality will very likely entail massive socio-economic change which will eventually contribute to raise the living standards of the total population in the developing countries. This may sound too theoretical and prohibitively expansive but in reality it may not be, provided national programs distribute the benefits of economic development broadly rather than enrich a microscopic few. In other words, the resource distribution should be done to favour heavily those at the lower rung of the economic ladder - the people who are usually the majority in the developing countries.

In general, the results of this analysis indicate that all the variables examined here, except infaltion, do play significant roles in discriminating among the three groups of countries - high living standard group, medium living standard group, and low living standard group. Much of the evidence of contributions of these variables towards such group differences is usually based on relatively simple, descriptive statistics, which may be misleading in several respects. The analysis in this paper has gone beyond simple description to uncover the evidence about the multivariate gruop differences and has identified functions which could predict well or serve as a reasonable description of the real world. As such, hopefully, it will provide more solid basis for discussions of social problems across countries.

References

Bennett, Spencer, and David Bowers, *An Introduction to Multivariate Techniques for Social and Behavioural Sciences,* John Wiley and Sons, New York, 1976.

Food Research and Action Group, *The Widening Gap: The Incidence and Distribution of Infant Mortality in the United States 1978~1982, Part I*, Washongton, D.C., Food Research and Action Center, 1984.

Hoffman and Hoffman Public Relations, *The International Human Suffering Index*, Population Crisis Committee, Washington, D.C., 1987.

Klecka, William R., *Discriminant Analysis,* Sage Publications Inc., Beverly Hills, 1980.

Lachenbruch, P. A., *Discriminant Analysis,* New York, Hafner, 1975.

Miller, C. A., "Infant Mortality in the U.S.", *Scientific American,* Vol.253, 1985, pp.31~37.

Morris, M. D., *Measuring the Condition of the World's Poor: The Physical Quality of Life Index.,* Oxford, Pergamon, 1979.

Newland, K., "Infant Mortality and the Health of Societies", *Worldwatch Paper* Vol.47, Washington, D.C., Worldwatch Institute, 1981.

Nie, Norman H., C.H. Hull, J. G. Jenkins, K. Steinbrenner and D. H. Bent, *Statistical Package for the Social Sciences,* Second Edition, McGraw Hill Book Company, 1975.