Available online at www.alphanumericjournal.com

alphanumeric journal

The Journal of Operations Research, Statistics, Econometrics and Management Information Systems

Volume 4, Issue 1, 2016

2016.04.01.STAT.01

ASSESSING THE EFFECTS OF VARIOUS SOCIO-ECONOMIC AND HEALTH INDICATORS ON HDI COUNTRY CATEGORIES

Özlem YORULMAZ*

Assist. Prof. Dr., Faculty of Economics, Istanbul University, Istanbul

Received: 27 January 2016

Accepted: 03 March 2016

Abstract

The human development indicator (HDI) is based on three indicators: standard of living, life expectancy, and education level. Although being widely known and commonly used, the accuracy of the HDI has been criticized in the literature due to the inadequacy of its indicators. The present study uses 11 indicators to classify countries and compares the results by country groups against similar HDI ranked country groups. Furthermore, using multinomial logistic regression analysis, the effects of the 11 indicators on the country categories of HDI are investigated. The findings show that although the main cluster characteristics are similar to HDI categories, some differences exist in the classification of countries. Health indicators have a striking effect on low HDI countries relative to high HDI countries. FDI inflows and CO2 emissions per capita are significant indicators for low and middle HDI relative to high HDI countries. However women's involvement in parliament and work are not distinctive or effective indicators.

Keywords: Cluster Analysis, Gender Effect, Health Indicators, Human Development Indicator, Multidimensional Scaling, Multinomial Logistic Regression

ÇEŞİTLİ SOSYO-EKONOMİK VE SAĞLIK GÖSTERGELERİNİN İNSANİ GELİŞME ENDEKSİ ÜLKE KATEGORİLERİ ÜZERİNDEKİ ETKİLERİNİN DEĞERLENDİRİLMESİ

Özet

İnsani Gelişme Endeksi (İGE) yaşam standardı, ortalama yaşam beklentisi ve eğitim düzeyi göstergelerine dayalıdır. İGE'nin tanınırlığı ve yaygın kullanımına karşın, endeksin yetersiz faktörlerle açıklanmasından dolayı doğruluğu konusu literatürde tartışılmaktadır. Bu çalışma 11 göstergeyi dikkate alarak ülkeleri kümelemiş ve elde edilen kümeleri İGE ülke kategorileri ile karşılaştırmıştır. Ayrıca söz konusu 11 faktörün İGE ülke kategorileri üzerindeki etkisi multinomial lojistik regresyon analizi ile değerlendirilmiştir. Elde edilen bulgular ülke sınıflarının genel özelliklerinin İGE kategorileriyle benzerlik gösterdiğini ama bununla beraber sınıflamada bazı farklılıkların olduğunu göstermiştir. Sağlık göstergeleri, düşük İGE kategorisinde yer alan ülkelerde, çok yüksek İGE kategorinde yer alan ülkelere oranla oldukça etkilidir. Yüksek IGE kategorisinde yer alan ülkeler referans olarak alındığında, doğrudan yabancı yatırımlar ve kişi başına düşen CO2 emisyonu göstergelerinin, düşük ve orta İGE kategorisinde yer alan ülkelerde istatistiksel olarak anlamlı olduğu görülmüştür. Bununla beraber, kadınların parlamentodaki ve iş gücündeki katılımları etkili ya da ayırt edici bir faktör değildir.

Anahtar Kelimeler : Insani Gelişme Endeksi, Sağlık Göstergeleri, Cinsiyet Etkisi, Multinomial Lojistik Regresyon Analizi, Kümeleme Analizi, Çok Boyutlu Ölçekleme Analizi

* yorulmaz@istanbul.edu.tr (Corresponding author)

1. INTRODUCTION

With respect to countries, the concept of development allows us to distinguish between rich and poor. It is related to "...shares of resources used to provide free health and education services, equitable distribution of income among social groups, effects of production and consumption on people's environment" (Soubbotina and Sheram, 2000). Gross domestic product (GDP) is an overly simplistic measure of development and the results in an inaccurate representation of actual human processes (Barreiro, 2006). However, for a long time, per capita income has been widely used as a means of making comparisons between the development indices of countries. If we look closer at countries with similar per capita GDP, it is possible to determine differences in the availability of clean air and water, conditions of education and health care systems, the unemployment ratio, the ratio of women's participation in parliament, and flows of foreign direct investment (FDI), among other factors. The objective of development is to create an enabling environment for people to enjoy long and creative lives (UNDP, 1990).

Researchers and policymakers have long investigated economic, social, political, and environmental indicators with which to profile countries, identify development policies and to rank their state of development. Before the 1990s, economic growth was commonly used as an indicator of development. Over time, economic growth was incorporated into multidimensional measures, that included several aspects of well-being, going beyond per capita income and GDP growth. Sen (1992) introduced the capability approach, which initiated the design of the human development index (HDI). However, term HDI did not appear in the literature until the publication of Human Development Report (1990), which at this point comprised a simple unweighted average of a nation's longevity, education, and income. Over time, modifications were made to the HDI. According to the Human Development Report published in 2014, the definition of HDI is a summary measure of achievements based on key dimensions of human development: a long and healthy life, access to knowledge, and a decent standard of living (Figure 1). Consequently, the HDI has become the geometric mean of normalized indices for each of these three dimensions.

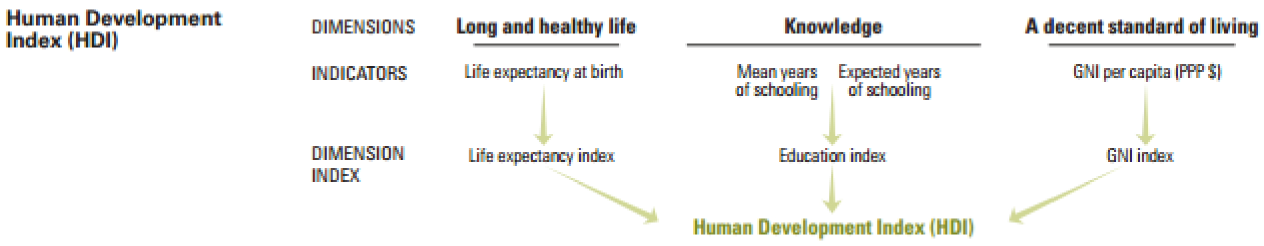

Figure 1. Sub-dimensions of HDI. Source: United Nations Report, 2014

The 2014 Human Development Report classified countries into four groups. The first group corresponds to "very high human development" with HDI values of 0.800-1.000. HDI values of 0.7-0.799 refer to "high human development" whereas values of 0.550-0.699 refer to "medium human development". The last group "low human development" has an HDI values falling below 0.55.

Despite its widespread use, the HDI has fallen under some criticism in the development literature. Most of these criticisms have focused upon the few arbitrary indicators comprising the HDI's sub-dimensions. Considering the data collection problems that some countries encounter, Srinivasan (1994) argues that "the HDI is conceptually weak and empirically unsound". Ranis et al. (2006) investigated the effects of 39 indicators and concluded that the HDI represented an incomplete measure. Wolff et al. (2010) argued that classifications based on the HDI were not reliable due to the presence of errors in the data, particularly in the health and education statistics.

The aim of this study is to investigate the effects of 11 socio-economic and health indicators on the HDI of country categories by way of multinomial logistic regression analysis. Additionally, this study examines the classification of countries in consideration of these indicators and evaluates the main characteristics of these groups by cluster analysis. The representation of HDI sub-dimension indicators is assessed based on the findings of the cluster analysis. To the best of our knowledge, this is the first study that has examined the effects of various indicators on HDI country categories. The results can be used to better inform policymakers on the shortcomings of countries ranked using the HDI and thus serve as a guide for decision making.

The remainder of the paper is organized into four sections. Section 2 reviews the relevant literature. Section 3 explains the methodology used throughout this study. Section 4 presents the data and findings, and Section 5 provides a conclusion.

2. LITERATURE REVIEW

Since the development of the HDI by the United Nations Development Programme in its 1990 report, the measure has been integrated in a wide range of studies. HDI embodies Sen's "capabilities" approach to understanding human well-being. Capabilities are instrumentalized in HDI as access to health, education, and goods (Stanton, 2007). However, as mentioned in the previous section, Ranis et al. (2006) investigated the effects of 31 indicators on human development and found HDI is an incomplete measure. In this section, we discuss several earlier studies with respect to the effects of socio-economic, demographic, and environmental variables on human development.

Lee et al. (1997) studied how HDI can be used to predict the infant and maternal mortality rate and reported HDI as a powerful predictor. Antony, Visweswarraro, and Balakrishna (1999) evaluated the representativeness of HDI for health and nutrition indicators of 174 countries. They found that dietary indicators are substantial for classification and that countries with high HDI have high rates of female education and lower rates of fertility and infant and maternal mortality. Self and Grabowski (2003) found that the contribution of health care expenditure as a percentage of GDP changes in developing and less developed countries. Ranis, Stewart and Somma (2005) found that under-five mortality rates are a good indicator of HDI. Similarly, Boutayeb and Serghihi (2006) showed that deficiencies of health (maternal and infant mortality rates) impede human development in the majority of Arab countries.

Sharma (1997) and Fukuda-Parr (2001) considered human development from a gender perspective and the importance of women on development level. Sharma (1997) argued that HDI could be improved by using sex-disaggregated data to examine the contributions various socio-economic factors.

Costa, Reybski, and Knopp (2011) found a positive relationship between per capita CO2 emissions for developing countries. Bedir and Yilmaz (2015) investigated the casual relationship between CO2 emissions and HDI for OECD countries and found a strong effect of CO2 on HDI in some of these countries.

Sharma, and Gani (2004) examined the effect of foreign direct investment (FDI) on HDI for middle- and low-income countries and found a positive effect. Later, Reiter and Steensma (2010) also found that FDI inflows affect the improvement of human development.

Saito (2003) emphasised the importance of education for human development. Njoh (2003) explored the relationship between urbanisation and the HDI, finding a positive correlation between the two for sub-Sahara Africa. Using discriminant analysis, Öztürk (2007) classified countries and predicted country categories through socioeconomic, education, and health indicators. He concluded that developed and undeveloped countries could be distinguished by their health indicators. The study also considered the proportion of seats in parliament, percentage of school enrolment and percentage of trade, concluding that these indicators could also be used to distinguish between developing and developed countries.

3. METHODOLOGIES

In this study, we employed cluster analysis to classify countries according to a range of socio-economic and health indicators. Multidimensional Scaling (MDS) was subsequently used to explore the links between the link between the indicators. In the final stage of analysis, we used multinomial logistic regression to examine the effects of the indicators on the HDI country categories.

3.1. Cluster Analysis

Cluster analysis comprises a range of methods used to identify homogeneous groups. As such, cluster analysis is used to ensure similarity within groups and, as much as possible, differences between groups. Three clustering techniques are defined in the literature: agglomerative hierarchical clustering, K-means, and density-based spatial clustering of applications with noise (DBSCAN). In our study, only hierarchical and K-means clustering were employed. In the K-means approach, the number of clusters is defined by researchers. At the beginning of the algorithm, K centroids are chosen and each observation is assigned to the closest centroid. The group of observations assigned to a cluster is called a cluster. With the assigned observations, the centroid of clusters is updated. The assignment procedure is repeated until the same centroids are obtained. The mentioned centroids are representative of each cluster's prototype. The objective functions and the centroids are outlined in Table 1.

Two approaches to hierarchical clustering are defined in the literature, namely, agglomerative and divisive. Compared to the divisive approach, the agglomerative procedure has received more attention in applications (Tan, Steinbach, & Kumar, 2005). The agglomerative approach begins with each observation as a singleton cluster and then, at each step, the two closest clusters are merged according to their proximities. The procedure continues until one cluster remains. Contrary to the agglomerative approach, divisive clustering begins with one cluster, which includes all observations. At each step, clusters are split until only singleton clusters of individual observations remain. The proximities used in agglomeration are calculated using different approaches, namely, single linkage (nearest neighbor), complete linkage (furthest neighbor), average linkage, centroid, and Ward's method (Sarstedt & Mooi, 2014).

Table 1. Objective Function and Centroid Choices in K-means Cluster Analysis

Proximity Function	Centroid	Objective Function
Manhattan (L_1)	Median	Minimize sum of L_1 distance of an object to its cluster centroid
Squared Euclidean (L_2)	Mean	Minimize sum of the squared L_2 distance of an object to its cluster centroid
Cosine	Mean	Maximize sum of the cosine similarity of an an object to its cluster centroid
Bregman divergence	Mean	Minimize sum of the Bregman divergence of an object to its cluster centroid

*Source: Tan, Steinbach, and Kumar (2005, p.501).

3.2. Multidimensional Scaling (MDS) Analysis

This technique determines the distances between objects and in order to provide a visual representation of objects in a low-dimensional space. Distances are referred to as proximities and proximity measures are described as dissimilarities, similarities, and correlations. Euclidean, Mahalanobis, quadratic Euclidean, Chebychev, Block, and Minkowski distances are defined in the literature. However, Euclidean distance has been preferred frequently in studies.

The Euclidean distance between the ith and jth points, d_{ij} is defined as

$d_{ij}^2 = \sum_{k=1}^{p} (x_{ik-} x_{jk})^2$ where p is the dimension of the observations.

The steps of the algorithm can be summarized as (Hintze, p.435):

1. Using Euclidean distances, elements of distance matrix *D* are calculated for observations.
2. Elements of matrix A, $A = \{-\frac{1}{2} d_{ij}^2\}$ are calculated by means of matrix *D*. Using matrix *A*, $B = \{a_{ij} - a_{i.} - a_{.j} + a_{..}\}$ is obtained where $a_{i.}$ is the average of all a_{ij} across *j*.The *p* largest eigenvalues $\lambda_1 > \lambda_2 > ... > \lambda_p$ and eigenvectors $L = (L_{(1)}, L_{(2)}, ..., L_{(p)})$ of matrix *B* are obtained and normalized so that $L_{(i)}^{'} L_{(i)} = \lambda_i$.
3. The rows of L refer to coordinates of the objects. $\hat{d}_{ij}$ distances from L are calculated.

To evaluate how well the dataset is represented by MDS, a goodness of fit statistic is defined for the actual values and their predicted values:

$$Stress = \sqrt{\frac{\sum\sum_{i<j} \left(d_{ij} - \hat{d}_{ij}\right)^2}{\sum\sum_{i<j} \left(d_{ij}\right)^2}}$$

The goodness of fit measure provides the closeness between original distances and predicted distances and is known as stress. Stress values < 0.05 define perfect fit.

3.3. Multinomial Logistic Regression

A multinomial logistic model determines the effects of explanatory variables on a subject chosen from a discrete set of options (Agresti, 2002). The result is a generalized binary model, the response variable having more than two categories.

Consider u possible outcomes {1,2,…,u} with multinomial probabilities $P[y = k] = p_k$, *k*=1,2,…,*u*. The probabilities can be parametrized as (Ledolter, 2013, p.132):

$$p_1 = P[y = 1] = \frac{1}{1 + \sum_{h=2}^{u} \exp(\alpha_h + x\beta_h)} \quad (1)$$

$$p_k = P[y = k] = \frac{\exp(\alpha_k + x\beta_k)}{1 + \sum_{h=2}^{u} \exp(\alpha_h + x\beta_h)} \quad \text{for } k=2,\dots,u.$$

The sum of probabilities is equal to 1. The interpretation of the odds-ratio is based on following equation"

$$\log\left(\frac{p_k}{p_1}\right) = \alpha_k + x\beta_k \quad \text{for } k=2,\dots,u. \quad (2)$$

The maximum likelihood estimation approach was used to estimate parameters. Multinomial logistic regression is similar to discriminant analysis. Discriminant analysis employs a regression line to separate sample into groups whereas multinomial logistic regression analysis uses probabilities and u-1 log odds equations to determine categories. Assumptions about multivariate normality and homoscedasticity are required for discriminant analysis; however, multinomial logistic assumes neither normality nor homoscedasticity

4. DATA AND FINDINGS

This study aimed to investigate the effects of selected socioeconomic and health indicators on HDI country classifications and country categories using data sourced from the 2013 United Nations Development Report. Data from 175 countries were considered as variables for the purposes of analysis, including: Maternal Mortality Rate (V1), Adolescent Birth Rate (V2), Women Share of Seats in Parliament (V3), Female Labor Force Participation Rate (V4), Deaths Due to Tuberculosis (V5), Increase in Rate of Physicians (V6), Public Health Expenditure (% of GDP) (V7), CO2 Emissions per Capita (V8), Exports and Imports (% of GDP) (V9), FDI Net Inflows (% of GDP) (V10), and Internet Use Rate (V11). Some potentially effective variables such as tertiary enrollment ratio, public expenditure on education (% of GDP), poverty headcount ratio at $3.10 per day, income share held by the highest 10%, income share held by the lowest 10%, and military expenditure (% of GDP) were excluded due to missing observations.

To consider standardized data, cluster analysis was employed. In the light of the findings from the dendogram from hierarchical clustering, we determined four clusters for K-means cluster analysis. It should be noted that the clusters containing countries are different from the four HDI categories (Appendix-I), particularly between the low and medium HDI groups. To determine the main characteristics of clusters, the mean scores displayed in Table 2 were used.

Table 2. Mean Scores of Indicators

Variables	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Maternal Mortality Rate	1.388	-0,26851	-0,635	-0,682
Adolescent Birth Rate	1.222	-0,03044	-0,846	-0,759
Share of Seats in Parliament (Women)	0.1048	-0,28347	-0,4915	0,4091
Female labor force participation rate	0.6726	-0,47464	-0.4661	0.0503
Deaths due to Tuberculosis	1.031	-0,18875	-0,5157	-0,506
Incrase rate of Physicians	-1.034	-0,32569	0,606	0.895
Public Health expenditure (% of GDP)	-0,255	-0,32016	-0,696	0,8337
CO2 Emissions per capita	-0,661	-0,35962	1.945	0,3896
Exports and Imports (% of GDP)	-0.2634	-0,22158	1.456	0,0327
FDI inflows (% of GDP)	0.2238	-0,06918	0,7678	-0.3564
Internet Use rate	-1.115	-0,30259	0,9781	1.108

As presented in Table 2, countries included in Cluster 1 have very low levels of Internet use rate and increase in rate of physicians. However, maternal mortality rate, adolescent birth rate, and deaths due to tuberculosis are at the highest levels. CO2 emissions per capita and percentage of exports and imports are at highest levels for Cluster3. Cluster 4 has the highest level of Internet use rate. It should be noted that the countries in the obtained clusters are different from the four HDI categories. As can be seen in Table 1 in the Appendix, the countries in Cluster 1 mainly correspond to the countries in the low HDI category, and countries in Cluster 4 mostly overlap with countries in the very high HDI category. However, Middle East and North Africa (MENA) countries included in the very high HDI category are not included in Cluster 4; instead, these countries constitute Cluster 3. Most of the countries in the high HDI category and some in the medium HDI country (e.g. South Africa, The Philippines, Indonesia, and India) are in Cluster 2. Although the main characteristics of these clusters are similar to the relevant HDI country categories, the same countries were not obtained with cluster analysis using the 11 indicators.

To provide insights about the effectiveness of indicators used to determine the classification of countries, MDS was employed. The visual representation considering the 11 indicators is presented in Figure 2. A stress value of 0.08 was obtained, which is an evidence of a “good” fit between the original distances and predicted distances on a two-dimensional plot. As can be seen from the positions of indicators, maternal mortality rate (V1) and exports and imports as percentage GDP (V9) are unrelated whereas variables V5, V6, V7, V8, and V10 have similar effects.

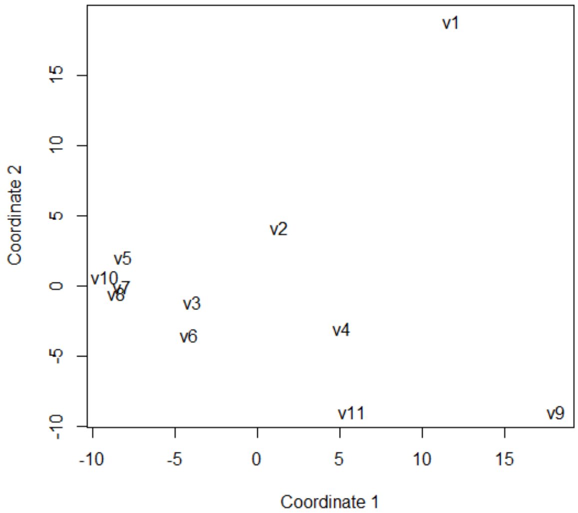

Figure 2. Perceptual Plot of Indicators

To assess the effects of indicators on country categories according to HDI and make a classification, discriminant analysis was employed. However, the assumptions of multivariate normality and homoscedasticity were not provided. [†] Therefore, multinomial logistic regression was used to investigate the effects of the indicators on HDI. Multinomial logistic regression is robust for revealing violations of assumptions of multivariate normality and covariance equality of groups. Furthermore, multinomial logistic regression does not assume a linear relationship between the dependent and independent variables (Akinci et al., 2007).

To determine the difference between the model without independent variables and the model with independent variables, a likelihood ratio Chi-Square test was performed and revealed that at least one of the regression coefficients in the model is not equal to zero (p=0.000). McFadden's pseudo R square for the model was 0.814. Since multinomial logistic model estimated k-1 [‡] equations, three equations are displayed in Table 3 with the high HDI country group chosen as the base outcome.

The first part of Table 3 displays indicators associated with very high scoring HDI countries. The rate of Internet use and percentage of exports and imports of GDP had significant effects for very high HDI countries relative to high HDI countries. With a one-unit increase in the indicator of percentage of exports and imports, the multinomial log odds for a country having a very high HDI relative to high HDI is expected to decrease 0.016 units while all the variables in the model remain constant. Considering the RRR statistic,[§] it can be interpreted that with a one-unit increase of the percentage of exports and imports variable, the relative risk of being in very high HDI country group is expected to decrease by a factor of 0.98 when the other variables are held constant. For the rate of Internet use variable, a one-unit increase is expected to increase the multinomial log odds by 1.38 units for the very high HDI category relative to the high HDI group, which means that with an increase of one unit in the rate of Internet use, the relative risk of being a very high HDI country is 1.38 times more likely. The second (middle) part of Table 3 corresponds to indicators associated with countries with medium HDI scores. Deaths due to tuberculosis, increase in rate of physicians, CO_2 emissions per capita, FDI net inflows as percentage of GDP, and Internet use rate were found to be significant variables that have effects on medium HDI countries relative to high HDI countries. With a one-unit increase of deaths due to tuberculosis, the multinomial log odds for medium HDI countries relative to high HDI countries is expected to increase 0.22 while the other variables in the model remain constant. Similarly, with a one-unit increase in increase in rate of physicians, the multinomial logs odds for medium HDI countries relative to high HDI countries is expected to increase by 0.16 units.

However, with separate one-unit increases of CO_2 emissions per capita, FDI net inflows as percentage of GDP, and Internet use rate, the relative risk of being in the medium HDI group relative to the high HDI group is expected to decrease by factors of 0.51, 0.76, and 0.89, respectively. The last part of Table 3 refers to the indicators associated with countries with low HDI scores. The significant variables and their effects on the model for medium HDI countries relative to high HDI countries are similar for the model with low HDI countries relative to high HDI countries. However, there is another significant indicator, namely, maternal mortality rate. A one-unit increase in the maternal mortality rate increases the relative risk of being in the low HDI group by a factor of 1.05.

Therefore, on the basis of the 11 socio-economic and health indicators, countries were classified into four groups. According to cluster means (Table 2), health related indicators (maternal mortality rate, deaths due to tuberculosis, and increase in rate of physicians) are effective for the classification of countries. Countries with high mean scores for health indicators are characteristically countries in the low HDI category. Alongside the health indicators, rate of Internet use, exports and imports as percentage of GDP, and CO_2 emissions per capita are also distinctive features for classification. Countries with high mean scores for the aforementioned indicators are characteristically countries in the high and very HDI categories. The perceptual map of indicators obtained from MDS analysis (Figure 2) presents the different patterns of maternal mortality rate from the remaining indicators. The findings of multinomial logistic regression (Table 3) indicate that maternal mortality rate and deaths due to tuberculosis are distinctive features of low HDI countries relative to high HDI countries. Rate of Internet use is a striking indicator that separates very high HDI countries from the remaining countries. For medium and low HDI countries relative to high HDI countries, variables of FDI inflow and CO_2 emissions per capita are decisive. Economic indicators, namely, exports and imports as percentage of GDP are important for classifying very high HDI countries relative to high HDI countries.

† Multivariate normality was examined with the Hz test and Royston test. For the homogeneity of covariances, Box-M tests were used.

‡ k is the number of levels of the dependent variable.

§ RRR refers to the relative risk ratio, which is calculated by exponentiating the multinomial logit coefficients.

Table 3. Parameter Estimates of Multinomial Logistic Regression

	Very High HDI				***Medium HDI***				***Low HDI***			
Base: High HDI	**Coef.**	**+Robust Std.Err.**	**Sig.**	**RRR***	**Coef.**	**Robust Std.Err.**	**Sig.**	**RRR***	**Coef.**	**+Robust Std.Err.**	**Sig.**	**RRR***
Maternal Mortality Rate	-0,0489	0,0517	0,3440	0,9522	0,0336	0,0221	0,1280	1,0342	0,0506	0,0238	0,03**	1,0519
Adolescent Birth Rate	-0,0417	0,0290	0,1510	0,9592	0,0483	0,0299	0,1060	1,0494	0,0497	0,0367	0,1770	1,0509
Share of seats in Parliament	0,1626	0,1474	0,2700	1,1765	-0,0192	0,0443	0,6640	0,9809	-0,0352	0,0760	0,6430	0,9654
Female labour forceparticipation rate	0,0098	0,0567	0,8620	1,0098	-0,0415	0,0429	0,3330	0,9594	-0,1023	0,0702	0,109	0,9028
Deaths due to Tuberculosis	-0,3995	0,2747	0,1460	0,6706	0,2247	0,1094	0,04**	1,2520	0,2488	0,1170	0,033**	1,2825
Physicians	-0,0338	0,4458	0,4480	0,9667	0,1647	0,0699	0,019**	1,1791	-0,1055	0,2969	0,7220	0,8999
Public health expenditure % of GDP	0,1251	0,3199	0,6960	1,1332	-0,4484	0,2939	0,1270	0,6386	0,1836	0,5381	0,7330	1,2015
CO2 emissions per capita	0,0102	0,0728	0,8080	1,0102	-0,6721	0,2973	0,02**	0,5106	-1,7090	0,8990	0,058***	0,1811
Exports and Imports % of GDP	-0,0159	0,0090	0,076***	0,9843	0,0363	0,0223	0,1030	1,0369	-0,0082	0,0492	0,8680	0,9918
FDI, inflows % of GDP	-0,1121	0,0699	0,1090	0,8940	-0,2703	0,1350	0,045**	0,7632	-0,3113	0,1492	0,037**	0,7325
Internet users rate	0,3245	0,1358	0,017**	1,3834	-0,1166	0,0462	0,012**	0,8899	-0,1932	0,0569	0,001*	0,8243
Intercept	-20,6781	8,4521	0,0140	-	-0,2052	1,9233	0,9150	-	1,4901	3,7356	0,6920	-

+ : Robust Standardize Errors
*, **, *** represent significance at 1, 5 and 10 % levels, respectively.

5. CONCLUSIONS

The HDI index has become one of the most widely used methods of ranking the development of countries according standards of living, life expectancy and literacy levels. Notwithstanding, the accuracy of the HDI has come under increased criticism in the literature due to a lack of representativeness and gaps in the data (Srinivasan, 1994; Ranis et al., 2006; Wolf et al., 2010).

The present study considers 11 socioeconomic and health developmental indicators for 175 countries, classifying them and assessing the effects of these indicators on the HDI country categories. We deduced four country clusters using cluster analysis, and used these country clusters as the basis of our HDI country categories. Comparison were made between the countries in the HDI categories and those in the clusters to reveal a number of startling differences. However, in evaluating the means scores for the clusters, we also revealed a number of commonalities with HDI categories. For instance, the rate of Internet use, exports and imports as percentage of GDP and CO_2 emissions per capita distinguish developed and undeveloped countries, with these indicators having higher mean scores for high and very high HDI countries. On the other hand, cluster mean scores for maternal mortality rate, deaths due to tuberculosis and increase in rate of physicians were greater for undeveloped countries, which were largely consistent with the low HDI countries. Notwithstanding, MDS analysis revealed that the maternal mortality rate was distinct from other indicators, as indicated by its position on the perceptual map.

Multinomial analysis was used to investigate the effect of the indicators on the HDI country categories; the findings alluding to similar interpretations as what were found with cluster analysis. Maternal mortality rate and deaths due to tuberculosis were the distinguishing features of low HDI countries relative to high HDI countries. Regarding the indicators of FDI inflow and CO_2 emissions per capita, these are more distinctive for medium and low HDI countries relative to high HDI countries. Specifically, if one of these indicators increases by just one unit, a country is much more likely to be in the high HDI category relative to the medium and low HDI categories. For classification in the very high HDI category relative to the high HDI category, exports and imports as a percentage of GDP is prominent. Specifically, a one unit increase in this indicator means that a country is much more likely to be in the high HDI group.

Additionally, the findings of this study reveal shortcomings in the classification of low income countries based on health issues and women's involvement in parliament and work. These turned out not to be particularly effective indicators for very high, medium, and low HDI countries relative to high HDI countries.

References

Agresti A., (2002). Categorical Data Analysis,John Wiley &Sons,New Jersey.

Akinci S., Kaynak E., Atilgan E.,Aksoy S.(2007). Where does the logistic regression analysis stand in marketing literature?A comparison of the Market Positioning of Prominent Journals, European Journal of Marketing. 41(5/6),537-567.

Antony GM, Visweswara Rao K, Balakrishna N.,(2001). Human Development Index: It's Suitability for the Assessment of Health and Nutritional Status of Indian States, EPW, Bombay, India, 36 ,p. 2976-9.

Barreiro I.C, (2006). Human Development assessment through the Human-Scale Development approach, Ph-D Thesis, Universitat Politecnica De Catalunya.

Bedir S., Yilmaz,V.M., (2015). CO2 Emissions and human development in OECD countries: granger causality analysis with a panel data approach, Eurasian Economic Review,p:1-14,doi 10.1007/s40822-015-0037-2

Boutayeb A., Serghini M.(2006),Heath indicators and human development in the Arab region, Int. J. Health Geogr,5(61), doi: 10.1186/1476-072X-5-61

Costa L., Rybski D., Kropp J.P., (2011). A Human Development Framework for CO2 Reductions. PLoS ONE 6(12): e29262. doi:10.1371/journal.pone.0029262

Fukuda-Parr, S.,(2001). Indicators of Human Development and Human Rights-Overlaps, Differences… and What about the Human Development Index? Statistical Journal of the United Nations Economic Commission for Europe Vol. 18., 239-248

Hintze,NCSS User's Guide I,NCSS, 2007 http://www.ncss.com/wpcontent/uploads/2012/09/NCSSUG1.pdf

Ledoleter, J. (2013),Data Mining and Business Analytics with R, John Wiley &Sons, New Jersey.

Lee K., Park, S., Knoshnood B., Hsieh KL., Mittendorf R., (1997).Human development index as a predictor of infant and maternal mortality rates. J.Pediatr., 131(3), 430-3.

Mooi E., and Sarstedt M., (2011).A Coincise Guide to Market Research, Springer-Verlag Berlin Heidelberg.

Njoh A.J., (2003). Urbanization and development in sub-Saharan Africa. Cities 20: 167–174

Öztürk, S.G., (2007). Classifying and predicting country types through development factors that influence economic, social, educational and health environments of countries,SWDI Proceedings papers S759, p.665-674.

Ranis, Stewart, Samman, (2006). Human Development: Beyond the Human Development Index, Journal of Human Development,7(3), p. 322-358.

Reiter S.L., Steensma H.K., (2010).Human Development and Foreign Direct Investment in Developing Counties: The Influence of FDI Policy and Corruption. World Development, 38(12), 1678-1691.

Saito M., (2003). Amartya Sen's Capability Approach to Education: A Critical Exploration. Journal of Philosophy of Education, 37: 17–33. doi: 10.1111/1467-9752.3701002

Self S., Grabowski R., (2003). How Effective is Public Health Expenditure in Improving Overall Health?A Cross-country Analysis, Applied Economics, 35,835-845.

Sen A., (1992). Inequality Reexamined, Russell Sage Foundation, New York.

Sharma B.,Gani A., (2004). "The Effects of Foreign Direct Investment on Human Development," Global Economy Journal, De Gruyter, 4(2), p. 1-20,

Sharma, S. D.,(1997). Making the Human Development Index (HDI) gender-sensitive. Gender and Development Vol. 5, No 1., 60—61

Soubbotina T.P., Sheram K.,(2000). Beyond Economic Growth, The World Bank.

Srinivasan T.N., (1994). Human Development: A New Paradigm or Reinvention of the Wheel?, American Economic Review, 84(2), p. 238-243.

Stanton E. ,(2007). The Human Development Index: A history,working paper,University of Massachussetts Amherst, http://scholarworks.umass.edu/cgi/viewcontent.cgi?article=1101&context=peri_workingpapers

Tan P.N., Steinbach M., Kumar V., (2005).Introduction to Data Mining, Pearson, New York

United Nations Development Programme (UNDP), Human Development Report (1990).

United Nations Development Programme (UNDP), Human Development Report (2014).

Wolff H., Chong H. and Auffhammer M., (2010). Classification, detection and consequences of data error: evidence from the human development index',

http://faculty.washington.edu/hgwolff/EJOnlineWebVersionofHDI_Wolffetal2010.pdf

Appendix

Country Classifications based on 11 indicators

Countries	HDI	Cluster	Countries	HDI	Cluster
Norway	1	4,000	Tunisia	2	2,000
Australia	1	4,000	Colombia	2	2,000
Switzerland	1	4,000	Jamaica	2	2,000
Denmark	1	4,000	Tonga	2	2,000
Netherlands	1	4,000	Belize	2	2,000
Germany	1	4,000	Maldives	2	3,000
Ireland	1	3,000	Samoa	2	2,000
United States	1	4,000	Botswana	3	2,000
Canada	1	4,000	Moldova (Republic of)	3	4,000
New Zealand	1	4,000	Egypt	3	2,000
Singapore	1	3,000	Turkmenistan	3	3,000
Hong Kong	1		Gabon	3	1,000
Sweden	1	4,000	Indonesia	3	2,000
United Kingdom	1	4,000	Paraguay	3	2,000
Iceland	1	4,000	Uzbekistan	3	2,000
Korea (Republic of)	1	4,000	Philippines	3	2,000
Israel	1	4,000	El Salvador	3	2,000
Luxembourg	1	3,000	South Africa	3	2,000
Japan	1	4,000	Viet Nam	3	2,000
Belgium	1	4,000	Bolivia	3	1,000
France	1	4,000	Kyrgyzstan	3	2,000
Austria	1	4,000	Iraq	3	2,000
Finland	1	4,000	Cabo Verde	3	2,000
Slovenia	1	4,000	Guyana	3	2,000
Spain	1	4,000	Nicaragua	3	2,000
Italy	1	4,000	Morocco	3	2,000
Czech Republic	1	4,000	Namibia	3	2,000
Greece	1	4,000	Guatemala	3	2,000
Estonia	1	3,000	Tajikistan	3	2,000
Brunei Darussalam	1		India	3	2,000
Cyprus	1	4,000	Honduras	3	2,000
Qatar	1	3,000	Bhutan	3	2,000
Slovakia	1	4,000	Timor-Leste	3	1,000
Poland	1	4,000	Syrian Arab Republic	3	2,000
Lithuania	1	4,000	Vanuatu	3	2,000
Malta	1	4,000	Congo	3	
Saudi Arabia	1	3,000	Kiribati	3	
Argentina	1	4,000	Equatorial Guinea	3	1,000
United Arab Emirates	1	3,000	Zambia	3	1,000
Chile	1	2,000	Ghana	3	1,000
Portugal	1	4,000	Bangladesh	3	1,000
Hungary	1	4,000	Cambodia	3	1,000
Bahrain	1	3,000	Sao Tome and Principe	3	2,000
Latvia	1	4,000	Kenya	4	1,000
Croatia	1	4,000	Nepal	4	1,000
Kuwait	1	3,000	Pakistan	4	2,000
Montenegro	1	2,000	Myanmar	4	
Belarus	2	4,000	Angola	4	1,000
Russian Federation	2	4,000	Swaziland	4	1,000
Oman	2	3,000	Tanzania	4	1,000
Romania	2	2,000	Nigeria	4	1,000
Uruguay	2	4,000	Cameroon	4	1,000
Kazakhstan	2	4,000	Madagascar	4	1,000
Barbados	2	4,000	Zimbabwe	4	

Countries	HDI	Cluster	Countries	HDI	Cluster
Bulgaria	2	4,000	Mauritania	4	1,000
Panama	2	2,000	Solomon Islands	4	2,000
Malaysia	2	3,000	Papua New Guinea	4	
Mauritius	2	2,000	Comoros	4	2,000
Seychelles	2		Yemen	4	2,000
Trinidad and Tobago	2	3,000	Lesotho	4	1,000
Serbia	2	4,000	Togo	4	1,000
Cuba	2		Haiti	4	
Lebanon	2	3,000	Rwanda	4	1,000
Costa Rica	2	2,000	Uganda	4	1,000
Iran	2	2,000	Benin	4	1,000
Venezuela	2	2,000	Sudan	4	2,000
Turkey	2	2,000	Djibouti	4	1,000
Sri Lanka	2	2,000	South Sudan	4	
Mexico	2	2,000	Senegal	4	1,000
Brazil	2	2,000	Afghanistan	4	1,000
Georgia	2	4,000	Côte d'Ivoire	4	1,000
Saint Kitts and Nevis	2		Malawi	4	1,000
Azerbaijan	2	4,000	Ethiopia	4	1,000
Grenada	2		Gambia	4	1,000
Jordan	2	2,000	Congo	4	1,000
Macedonia	2	4,000	Liberia	4	1,000
Ukraine	2	4,000	Guinea-Bissau	4	
Algeria	2	2,000	Mali	4	1,000
Peru	2	2,000	Mozambique	4	1,000
Albania	2	2,000	Sierra Leone	4	1,000
Armenia	2	2,000	Guinea	4	1,000
Bosnia and Her.	2	4,000	Burkina Faso	4	1,000
Ecuador	2	2,000	Burundi	4	1,000
China	2	2,000	Chad	4	1,000
Fiji	2	2,000	Eritrea	4	1,000
Thailand	2	2,000	Central African Republic	4	1,000
Dominica	2		Niger	4	1,000
Libya	2	2,000			