Department of Statistics

Human Development Index and Human Poverty Index for Indian states, 2005: Multivariate Statistical Analysis of basic indicators

Master Thesis

(15 ECTS)

Spring 2010

Submitted by: Ismail Shah & Vinod Gosavi

Supervisor: Prof. Björn Holmquist

Department of Statistics
School of Economics and Management, Lund University, Sweden

Abstract

Human development index (HDI) is a summary measure used around the world that indicates whether a country is developed, still developing, or undeveloped. HDI incorporates the main factors of human life such as health, education and income. Deprivation in these areas of human life is measured by calculating human poverty index (HPI). The purpose of this thesis is to find the human development index and human poverty index for all states of India and perform multivariate statistical analysis on the indicators used in the calculation of HDI and HPI.

We calculate HDI and HPI for all 33 states of India and observe variation among states according to HDI and HPI. Ranking these 33 states helps us to compare the states according to both HDI and HPI. Multivariate techniques are used to analyze the main indicators used in the calculation of HDI and HPI. Performing factor analysis separately for the indicators used in the calculation of HDI and those of HPI, we find that two factors explain 85% and 66% variation in the data respectively. We observe that indicators used in HDI calculation are strongly correlated to each other but we don't found any statistically significant correlation among the indicators used in the calculation of HPI. Using cluster analysis for 7 indicators from HDI and HPI, we divide all the 33 states into two clusters size of 13 and 20 each. We observe that 12 states in cluster 1 are among the 14 top ranks states according to HDI. We observe that most of the states are highly developed and rich in 1^{st} cluster. The ANOVA table for these 7 indicators shows that the indicators from HPI are not statistically significant. After removing these indicators from our analysis we obtain two clusters again of same size and having same states as in our first cluster analysis. We conclude that to divide these 33 states into two clusters we only need the indicators from HDI.

Contents

1. Introduction

Gross domestic product per capita (GDP per capita) is an indicator of the average standard of living of individual members of the population. An increase in GDP per capita signifies national economic growth. Before 1990, GDP per capita and GDP as a whole were used to evaluate the human development or country development and for comparison purpose between countries. Despite this importance of GDP per capita, economic theorist finds flaws and problem with this indicator since this indicator only cover the financial side of human development. In 1990 an economist Mahbub ul Haq proposed a new method for evaluation of human development which is named as human development index (HDI). The concept of HDI occupies great importance since it covers both economic and social factors of human development. The HDI is now mostly used to evaluate human development in a state or country and for comparison purpose among states or countries.

The human development index takes into account three basic dimensions of human development, namely, longevity, knowledge and decent standard of living. Longevity is measured by life expectancy at birth, knowledge is measured by a combination of the adult literacy rate and the combined primary, secondary, and tertiary gross enrolment ratio and standard of living by GDP per capita (PPP US$). Here PPP stand for purchasing power parity and is a criterion for an appropriate exchange rate between currencies when a representative basket of goods in two different countries cost the same.

The three indices are calculated for longevity, knowledge and decent standard of living and HDI is calculated as average of these three indices.

The human poverty index for developing countries measures human deprivations in the three dimensions of human development as HDI i.e. longevity, knowledge and a decent standard of living. Deprivation in longevity is measured by calculating the percentage of people not expected to survive to age 40 years; deprivation in knowledge is measured by the percentage of adults who are illiterate; deprivation in a decent standard of living is measured by three variables: the percentage of people not having sustainable access to safe drinking water source; the percentage of people

without access to health services and the percentage of children below the age of five who are underweight. Human poverty index for selected high-income OECD countries includes social exclusion, in addition to the three dimensions in HPI for developing countries.

India is officially known as Republic of India and comprises of 28 states and seven union territories. India is the second largest populated country in the world and world's largest democracy. India is divided into six major zones East India, West India, North India, South India, Northeast India and Central India. The total geographical area of India is 3,287,240 square kilometers (1,269,210.5 sq miles). As per census data 2001, the total population of India is 1,028,737,436. The information about population of states and union territories is given in appendix II. Every state in India has several districts. The number of districts in each state varies widely. For instance Uttar Pradesh has 71 districts while Goa has just 2 districts. Irrespective of the size of the state or the union territory, every state or union territory is diverse in the demography, dress, festivals, culture, history and languages etc which makes it diverse. More details about India can be found on Wikipedia.

The first objective of this thesis is to calculate the Human Development Index and Human Poverty Index for 33 states of India. The second objective is to perform multivariate statistical analysis on indicators used to calculate HDI and HPI. In multivariate statistical analysis, factor analysis is use to identify those indicators which have high influence on HDI and HPI and cluster analysis is use to separate 33 states of India into different groups. SPSS version 18 is used for all the analysis.

2. Data: India Human Development Survey (IHDS), 2005

The data sets used in this thesis are from India Human Development Survey (IHDS), 2005. The India Human Development Survey 2005 (IHDS) is a nationally representative, multi-topic survey of 41,554 households in 1,503 villages and 971 urban neighborhoods across India. The principal investigators of IHDS, 2005 are Sonalde Desai, University of Maryland, Reeve Vanneman, University of Maryland and National Council of Applied Economic Research, New Delhi. The data of India Human Development Survey (IHDS) 2005 is obtained from Inter-university Consortium for Political and Social Research, University of Michigan.

The India Human Development Survey, 2005 has eight data sets consisting information on different aspect of human development, denoted by DS1, DS2, DS3, DS4, DS5, DS6, DS7 and DS8. The data set DS1 contains information of 215754 individual on 211 variables. The data set DS2 contains information on 41554 household on 211 variables. The data set DS3 contains information on medical facility on 130 variables. The data set DS4 contains information on nonresident on 125 variables. The data set DS5 contains information about primary school on 102 variables. The data set DS6 contains information about birth history on 32 variables. The data set DS7 contains information about 1501 villages on 378 variables. The data set DS8 contains information about crops on 21 variables.

For our study purpose, we use the INCOME variable from DS1 consisting information about income for each individual. Other variables are Ro5, which consists of observations on age of each individual, ED2 which consists of observations on literacy of each individual, ED4 which consists of observations on whether the individual is presently enrolled for education. We also uses the variable AP4 being observations on weight of individuals.

From DS2, we consider the variables DE4A, DE4B, DE4C consisting of data regarding deaths in the families. We also use the variable WA7 consisting information about drinking water facilities.

From DS7, we use the variables VMA1- VMA13 containing information regarding the access to health services.

3. Methods, Calculations and Results for HDI and HPI

This section contains the methodology for calculating HDI and HPI. In this section we start with the method for calculating HDI and explain how to calculate different indices used in the calculation of HDI. The calculations are exemplified by calculating HDI for India. We then discuss methodology for calculating different indicators used in the calculation of HPI. Again these calculations are exemplified by calculating HPI for India. In the last section we discuss the results obtained for all 33 states of India. Different measures for explaining variation in poverty and development are discussed in Antony and Rao (2007).

3.1 Method for calculating HDI

Human development index (HDI) is a composite index which combines economic and social factors to evaluate the development of a person, state or a country. There are three indices which are combined to give HDI and there are three steps to create this index.

- Describing dimensions for HDI to include.
- Describing indicators that capture these dimensions.
- Describing the method for combining these indicators to get a single value which represents HDI.

A long and healthy life (Health index), being knowledgeable (Education index) and a decent standard of living (Income index) are the three dimensions proposed by United Nation Development Programme (UNDP) for calculating HDI. In our analysis we use some specified indicators describe by UNDP which capture these three dimensions. These indicators are then normalized between 0 and 1 by constructing an index for each dimension. To create an index we use the following general formula.

$$\text{Index} = \frac{\text{Actual value of X} - \text{Min of X}}{\text{Max of X} - \text{Min of X}}$$

where

Actual value of X = value calculated from data for indicator X.

Min of X = minimum value that indicator *X* can obtain.

Max of X = maximum value that indicator X can obtain.

The minimum and maximum values for different indicators are proposed by UNDP and these are mentioned in the following text.

Health index

Life expectancy at birth is used as an indicator for long and healthy life. For calculating the health index we use the general formula. The UNDP use 25 and 85 years as goal posts for health index. i.e.

$$\text{Health Index} = \frac{\text{Life Expectancy} - 25}{85 - 25}$$

The life expectancy calculation method is described in Appendix I.

Education index

The education index is an aggregate index derived from two indices. One is the adult literacy rate i.e. the literacy rate of the population of age 15 years and above. The other is the gross enrolment ratio which is obtained by combining primary, secondary and tertiary gross enrolment in age 6 to 14 years. These two indices are combined, giving two third weights to adult literacy and one third weights to gross enrolment, to obtain the education index. The following mathematical form is thus used to obtain the education index.

$$\text{Education Index} = \frac{2}{3} * \text{Adult literacy index} + \frac{1}{3} * \text{Gross enrolment index}$$

To calculate adult literacy ratio (ALR) index and gross enrolment ratio (GER) index, UNDP uses 0% and 100% values for minimum and maximum goal posts in these formulas. Hence

$$\text{ALR Index} = \frac{\text{Adult literacy percentage } - 0}{100 - 0}$$

and

$$\text{GER Index} = \frac{\text{Gross enrolment percentage} - 0}{100 - 0}$$

Income index

To calculate income index, UNDP uses GDP per capita (PPP$). The UNDP uses 'adjusted per capita income for countries' to calculate income index. The GDP per capita (PPP$) is then transformed to log (GDP per capita (PPP$)) because increases of income at lower levels have then greater impact on the income index. The UNDP uses minimum and maximum goal posts 100$ and 40000$ for the income index. The following formula is thus used to calculate the income index.

$$\text{Income Index} = \frac{\log(\text{GDP per capita(PPP\$)}) - \log(100)}{\log(40000) - \log(100)}$$

The GDP per capita in Indian rupees were calculated for each states of India and then multiplied by the average nominal exchange rate of US $(2005) to obtain the GDP per capita in dollars. This is not exactly the same as GDP per capita (PPP$). This can be called as GDP per capita converted to dollars. This will affect income index values but the variation among the values of income index for different states will be the same since this effect is constant for all 33 states. Income index is then calculated for each states of India according to the following formula.

$$\text{Income index} = \frac{\log(\text{GDP per capita(\$)}) - \log(100)}{\log(40000) - \log(100)}$$

We are also aware of this fact that costs of living in different states are different. But states with high cost of living also provide more opportunity for a person and mostly earning is also high in that state. This fact minimizes the difference of cost of living among different states. Also in calculation of income index we use logarithm (log) for values used in the formula that reduce the influence of higher values.

Human Development Index

After computing above three indices HDI is calculated as simple average of these indices for each state of India. i.e.

$$\text{HDI} = \frac{\text{Health index} + \text{Education index} + \text{Income index}}{3}$$

Calculation of HDI for India

1. Life Expectancy Index

The life expectancy at birth for India is calculated by using variable Ro5 which contain information on age of 215754 individual from data set DS1. Other variables used are DE4A, DE4B and DE4C containing information about age on deaths in families. Using these variables, we calculate life expectancy at birth as per methodology describes in appendix I and obtain 64.07 years for India. Therefore

$$\text{Life expectancy Index} = \frac{64.07-25}{85-25} = 0.65$$

2. Education Index

To calculate education index, we first calculate the Adult Literacy Ratio index and gross enrolment ratio index for India. We have total number of individual above age 15 is 147292 and from these individual 97925 are literate. Therefore, ALR = 0.66
Also the total numbers of individual between age group 6 to 14 is 39974 and from these individual 37874 are enrolled in school. Therefore, GER = 0.95. Hence

$$\text{Education index} = 2/3\ \text{ALR} + 1/3\ \text{GER} = 2/3*0.66 + 1/3*0.95 = 0.76$$

3. Income Index

We calculate GDP per capita ($) to calculate income index. For this purpose first GDP (Rupees) as a whole is calculated and then divided by the total number of population. Then GDP per capita (Rupees) is converted to GDP per capita ($) according to average nominal exchange rate of 2005. Therefore for India we obtain GDP per capita ($) is 236.06. Hence

$$\text{Income index} = \frac{\log(236.06)-\log(100)}{\log(40000)-\log(100)} = 0.14$$

Human Development Index

Finally to obtain Human development index we calculate an average of these three indices.

$$\text{HDI} = \frac{0.65+0.76+0.14}{3} = 0.52$$

3.2 Method for calculating HPI

The Human Poverty Index (HPI) measures deprivations in three dimensions of human development. These three dimensions of human life consist of longevity, knowledge and a decent standard of living. Deprivation in long and healthy life is denoted by P_1 and calculated by the percentage of people not expected to survive to age 40. Lack of knowledge is denoted by P_2 and calculated as the percentage of adults who are illiterate. Deprivation in decent standard of living, $P_{3,}$ is a composite measure in three dimensions of decent standard of living. These three dimensions are

- $P3_1$ which is calculated as percentage of people without access to safe drinking water.
- $P3_2$ which is calculated as percentage of moderately and severely underweight children under 5 year's age.
- $P3_3$ is the percentage of people without access to health services.

Then these three dimensions are combined in the following way to get $P_{3.}$

$$P_3 = \frac{P3_1 + P3_2 + P3_3}{3}$$

We calculate these all indicators from our data for all 33 states of India except $P3_2$. In our analysis we use a constant value for $P3_2$ as we were facing a large number of missing values for this indicator in the data set. We calculate this indicator from the available data and use this constant measure for every state of India. Finally Human poverty index is calculated from the following formula.

$$\text{HPI} = [{}^{1}/_{3}\,({P_1}^3 + {P_2}^3 + {P_3}^3)]^{1/3}$$

Calculation of HPI for India

We calculate all the indicators for HPI. The following are the results we obtained from our calculations.

P_1 = percentage of people not expected to survive to age 40 years = 0.13

P_2 = percentage of adults who are illiterate = 0.34

$P_3 = (P3_1 + P3_2 + P3_3)/3 = (0.08+0.43+0.17)/3 = 0.22$

where

$P3_1$ = percentage of people without access to safe drinking water = 0.08

$P3_2$ = percentage of moderately and severely underweight children < 5 years age= 0.43

$P3_3$ = percentage of people without access to health services = 0.17

Finally we calculate HPI for India.

$HPI = [1/3(P_1^3 + P_2^3 + P_3^3]^{1/3} = 0.26$

3.3 Results

For each state of India the index values for the three different components of human development is obtained as per methodology discussed in this chapter. These three indices are Education, Health, and Income. From these three indices HDI is obtained for all states of India. We also calculate all the indicators of HPI for all states of India and then calculate HPI for each state.

Table 1 contains education index, health index, income index and HDI for all states of India. The states are ranked according to the size magnitude of HDI. From this table we can observe that Chandigarh is on first rank and the HDI value for this state is 0.6989. Meghalaya is on 33rd rank and the HDI value for this state is 0.4177. The union territories Chandigarh, Delhi and Pondicherry are among the first ten states according to HDI.

Table 2 gives values of P_1, P_2, P_3 and HPI for all states of India. The states are also ranked according to HPI. Here we can see that Kerala is on first rank and the HPI value for this state is 0.1115. Deprivation in human development is high in Rajasthan and the HPI value for this state is 0.3408.

The different indicators used to calculate HDI and HPI for all states of India are shown in Table 3 and Table 4. Multivariate analysis for all these indicators is performed in section 4.

Table 1: Education, Health and Income indices along with HDI for 33 states of India

Rank	Name of State	Education Index	Health Index	Income Index	Human Development Index
1	Chandigarh	0.9271	0.8742	0.2954	0.6989
2	Mizoram	0.9828	0.7949	0.3089	0.6955
3	Nagaland	0.9516	0.8851	0.2170	0.6846
4	Goa	0.9050	0.8542	0.2392	0.6662
5	Pondicherry	0.8516	0.8759	0.2178	0.6484
6	Kerala	0.9536	0.7724	0.1936	0.6398
7	Delhi	0.8759	0.7674	0.2371	0.6268
8	Arunachal Pradesh	0.8099	0.7619	0.2180	0.5966
9	Punjab	0.7864	0.8200	0.1746	0.5937
10	Jammu & Kashmir	0.7448	0.8411	0.1927	0.5929
11	Himachal Pradesh	0.8342	0.7389	0.1895	0.5875
12	Maharashtra	0.8164	0.7490	0.1535	0.5730
13	Gujarat	0.7833	0.7664	0.1661	0.5719
14	Sikkim	0.8457	0.5968	0.2629	0.5684
15	Uttaranchal	0.7771	0.6917	0.1434	0.5374
16	Assam	0.8238	0.6354	0.1415	0.5336
17	West Bengal	0.7834	0.6363	0.1695	0.5297
18	Haryana	0.7284	0.6912	0.1690	0.5295
19	Jharkhand	0.7414	0.6967	0.1420	0.5267
20	Tripura	0.8364	0.5129	0.2160	0.5218
21	Karnataka	0.7628	0.6405	0.1452	0.5162
22	Manipur	0.8857	0.3798	0.2721	0.5126
23	Rajasthan	0.6630	0.7440	0.1225	0.5098
24	Madhya Pradesh	0.6928	0.7338	0.0558	0.4941
25	Tamil Nadu	0.7875	0.5407	0.1470	0.4917
26	Dadra and Nagar Haveli	0.7252	0.6327	0.1076	0.4885
27	Chhattisgarh	0.7089	0.6166	0.0929	0.4728
28	Andhra Pradesh	0.6748	0.5885	0.1412	0.4682
29	Orissa	0.7360	0.5753	0.0694	0.4602
30	Daman and Diu	0.7746	0.3446	0.1466	0.4219
31	Bihar	0.6892	0.5093	0.0589	0.4191
32	Uttar Pradesh	0.6878	0.4783	0.0874	0.4178
33	Meghalaya	0.8607	0.1817	0.2107	0.4177

Table 2: P_1, P_2, P_3 and HPI for 33 states of India

Rank	Name of State	Percentage of people not expected to survive to age 40 yrs	Percentage of Adults who are illiterate	Percentage of people deprived from decent standard of living	Human Poverty Index
1	Kerala	0.0554	0.0688	0.1542	0.1115
2	Nagaland	0.1149	0.0692	0.2114	0.1556
3	Delhi	0.0337	0.1707	0.2235	0.1754
4	Himachal Pradesh	0.0941	0.2386	0.1604	0.1835
5	Tripura	0.1817	0.2072	0.1972	0.1959
6	Assam	0.1285	0.2109	0.2238	0.1964
7	Goa	0.0000	0.1376	0.2740	0.1977
8	Pondicherry	0.0000	0.2166	0.2528	0.2062
9	Chandigarh	0.0000	0.1018	0.2945	0.2070
10	Maharashtra	0.0746	0.2565	0.2193	0.2101
11	West Bengal	0.0977	0.2846	0.1570	0.2102
12	Sikkim	0.0000	0.2202	0.2620	0.2122
13	Gujarat	0.0978	0.2892	0.1840	0.2186
14	Mizoram	0.0000	0.0129	0.3222	0.2234
15	Punjab	0.0935	0.3033	0.2033	0.2313
16	Meghalaya	0.2616	0.1603	0.2622	0.2372
17	Arunachal Pradesh	0.0000	0.2756	0.2762	0.2410
18	Tamil Nadu	0.1924	0.3012	0.2031	0.2426
19	Uttaranchal	0.0645	0.3111	0.2799	0.2593
20	Jharkhand	0.0957	0.3525	0.2282	0.2662
21	Daman and Diu	0.1695	0.3281	0.2552	0.2666
22	Dadra and Nagar Haveli	0.1266	0.3738	0.1417	0.2670
23	Orissa	0.1345	0.3570	0.2285	0.2712
24	Karnataka	0.1170	0.3310	0.2835	0.2724
25	Haryana	0.1513	0.3897	0.1622	0.2815
26	Chhattisgarh	0.0993	0.4065	0.2097	0.2955
27	Jammu & Kashmir	0.0428	0.3727	0.3100	0.3008
28	Manipur	0.3819	0.1484	0.2968	0.3050
29	Madhya Pradesh	0.1054	0.4353	0.2246	0.3163
30	Bihar	0.2060	0.4317	0.2009	0.3189
31	Andhra Pradesh	0.1265	0.4588	0.2037	0.3292
32	Uttar Pradesh	0.2580	0.4379	0.2079	0.3323
33	Rajasthan	0.0672	0.4753	0.2225	0.3408

Table 3: Indicators used for HDI

Name of State	Adult Literacy	Gross Enrollment	Life Expectancy	Per Capita Income(Rs.)
Chandigarh	0.8982	0.9848	77.4529	25800.2611
Mizoram	0.9871	0.9740	72.6924	27970.4785
Nagaland	0.9308	0.9933	78.1035	16130.8238
Goa	0.8624	0.9902	76.2550	18428.4356
Pondicherry	0.7834	0.9880	77.5532	16209.8937
Kerala	0.9312	0.9982	71.3442	14019.8514
Delhi	0.8293	0.9691	71.0466	18190.4213
Arunachal Pradesh	0.7244	0.9810	70.7143	16232.1984
Punjab	0.6967	0.9657	74.2024	12511.0999
Jammu & Kashmir	0.6273	0.9798	75.4660	13948.4565
Himachal Pradesh	0.7614	0.9797	69.3345	13681.6130
Maharashtra	0.7435	0.9620	69.9397	11027.0549
Gujarat	0.7108	0.9281	70.9856	11887.9087
Sikkim	0.7798	0.9773	60.8060	21235.9145
Uttaranchal	0.6889	0.9534	66.5026	10379.0880
Assam	0.7891	0.8932	63.1216	10263.8059
West Bengal	0.7154	0.9193	63.1764	12139.0160
Haryana	0.6103	0.9646	66.4709	12102.9538
Jharkhand	0.6475	0.9293	66.8008	10293.8622
Tripura	0.7928	0.9236	55.7725	16033.4639
Karnataka	0.6690	0.9505	63.4311	10491.2615
Manipur	0.8516	0.9540	47.7887	22441.2151
Rajasthan	0.5247	0.9396	69.6396	9156.3344
Madhya Pradesh	0.5647	0.9488	69.0260	6139.8849
Tamil Nadu	0.6988	0.9648	57.4448	10603.0917
Dadra and Nagar Haveli	0.6262	0.9231	62.9632	8374.1193
Chhattisgarh	0.5935	0.9399	61.9986	7670.7539
Andhra Pradesh	0.5412	0.9421	60.3100	10244.2797
Orissa	0.6430	0.9220	59.5182	6663.1233
Daman and Diu	0.6719	0.9800	45.6780	10578.3384
Bihar	0.5683	0.9308	55.5581	6256.3861
Uttar Pradesh	0.5621	0.9391	53.6991	7421.2769
Meghalaya	0.8397	0.9028	35.9028	15533.4673

Table 4: Indicators used for Human Poverty Index

Name of State	Percentage of people not expected to survive to age 40 yrs	Adult illiteracy	Percentage of people without access to safe drinking water	Percentage of people without access to health services	Percentage of underweight children under 5 yrs of age
Kerala	0.0554	0.0688	0.0374	0.0000	0.4251
Nagaland	0.1149	0.0692	0.0091	0.2000	0.4251
Delhi	0.0337	0.1707	0.0454	0.2000	0.4251
Himachal Pradesh	0.0941	0.2386	0.0544	0.0017	0.4251
Tripura	0.1817	0.2072	0.0237	0.1429	0.4251
Assam	0.1285	0.2109	0.0520	0.1944	0.4251
Goa	0.0000	0.1376	0.0637	0.3333	0.4251
Pondicherry	0.0000	0.2166	0.0000	0.3333	0.4251
Chandigarh	0.0000	0.1018	0.4584	0.0000	0.4251
Maharashtra	0.0746	0.2565	0.1101	0.1228	0.4251
West Bengal	0.0977	0.2846	0.0142	0.0317	0.4251
Sikkim	0.0000	0.2202	0.0277	0.3333	0.4251
Gujarat	0.0978	0.2892	0.0646	0.0625	0.4251
Mizoram	0.0000	0.0129	0.5414	0.0000	0.4251
Punjab	0.0935	0.3033	0.1348	0.0500	0.4251
Meghalaya	0.2616	0.1603	0.1949	0.1667	0.4251
Arunachal Pradesh	0.0000	0.2756	0.0702	0.3333	0.4251
Tamil Nadu	0.1924	0.3012	0.0593	0.1250	0.4251
Uttaranchal	0.0645	0.3111	0.0813	0.3333	0.4251
Jharkhand	0.0957	0.3525	0.0673	0.1923	0.4251
Daman and Diu	0.1695	0.3281	0.0074	0.3333	0.4251
Dadra and Nagar Haveli	0.1266	0.3738	0.0000	0.0000	0.4251
Orissa	0.1345	0.3570	0.0315	0.2289	0.4251
Karnataka	0.1170	0.3310	0.1931	0.2324	0.4251
Haryana	0.1513	0.3897	0.0486	0.0128	0.4251
Chhattisgarh	0.0993	0.4065	0.0338	0.1702	0.4251
Jammu & Kashmir	0.0428	0.3727	0.1548	0.3500	0.4251
Manipur	0.3819	0.1484	0.1321	0.3333	0.4251
Madhya Pradesh	0.1054	0.4353	0.0724	0.1765	0.4251
Bihar	0.2060	0.4317	0.0301	0.1475	0.4251
Andhra Pradesh	0.1265	0.4588	0.0689	0.1170	0.4251
Uttar Pradesh	0.2580	0.4379	0.0119	0.1866	0.4251
Rajasthan	0.0672	0.4753	0.1046	0.1379	0.4251

4. Multivariate analysis of indicators

This section describes multivariate techniques to analyze the different indicators used in the calculation of HDI and HPI. Initially factor analysis is performed for the indicators used in the calculation of HDI and also for those used in HPI. With help of cluster analysis using different indicators used in the calculation of HDI and HPI all the 33 states are divided into two groups.

4.1 Factor analysis

Factor analysis was first introduced nearly 100 years ago by psychologist Charles Spearman to define and measure intelligence. The main purpose of factor analysis is to identify a few underlying, but unobservable, random quantities called factors that explain the pattern of correlation within a set of observed variables. In large data sets we often use factor analysis for data reduction to identify a small number of factors that explain most of the variance and covariance in the data set. Factor analysis try to identify such groups that within the group all the variables are highly correlated but relatively independent or less correlated with variables in other groups. Each group of variables represents a single underlying structure or factor which is responsible for the observed correlation within the variables in that group. More details can be found in Johnson and Wichern (2007).

4.1.1 Factor analysis methods

In factor analysis the two most popular methods of parameter estimation are the principal component analysis (and the related principal factor) method and the maximum likelihood estimation method. In our analysis we use principle component analysis as this analysis technique allow us for the extraction of as many significant factors as possible from our data set. This method explains the covariance structure in terms of just a few common factors.

4.1.2 Calculations and Results

In the first factor analysis we use four indicators; Gross enrolment (GE), Adult literacy (AL), Life expectancy (LE), Per capita income (PCI). These indicators are those occurring in calculation of HDI. To reduce the influence of extreme values of the indicators, they were standardized before used in factor analysis. Standardization

of variables is very useful technique and provides a lot of simplification. It reduces the influence of high values of variables on the results. By standardizing all the values of a variable are scaled in between a smaller range and all the values are laying around their mean value. Also it is easier to handle variables for which the mean is zero and for which the variance is one. Standardization is a simple procedure in which we subtract each value from its mean and divide by its standard deviation.

In Table 5 below, all 4 components (factors) would be needed to explain 100% of the variance in the data. The eigenvalues measure the amount of variation in the total sample accounted for by each factor. Since all the variables were standardized, we can use the conventional criterion of stopping when the initial eigenvalue drops below 1.0. Here only 2 of the 4 factors were actually extracted in this analysis. These two account for 85% of the variance in the data. The extracted two factors have the following form.

Factor 1 = 0.77ZGE + 0.58ZLE + 0.86ZPCI + 0.83ZAL
Factor 2 = 0.39ZGE + 0.72ZLE - 0.40ZPCI - 0.45ZAL

From Factor 1 above we can see that ZPCI and ZAL are contributing maximum in the calculation of this factor. While in the calculation of Factor2 ZLE is contributing maximum. These results also show the correlation structure between these variables.

Table 5: Extraction of factors

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.370	59.246	59.246	2.370	59.246	59.246
2	1.033	25.827	85.073	1.033	25.827	85.073
3	.416	10.404	95.477			
4	.181	4.523	100.000			

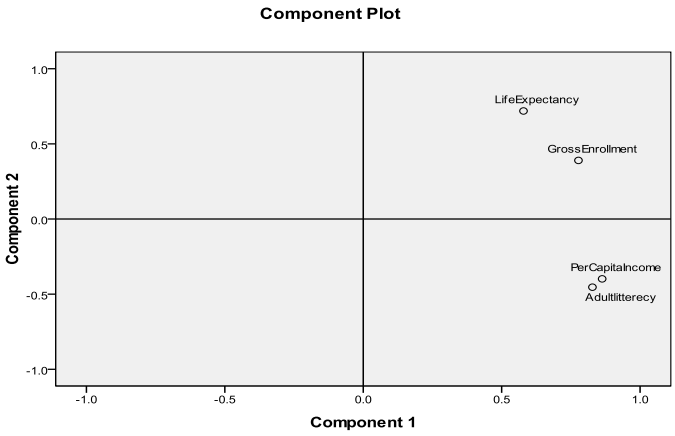
Table 6 shows the correlation coefficients among GE, LE, PCI and AL. Here we can see that correlation coefficient between PCI and AL is very high, Pearson's r = 0.811 (p=0.000). This indicates a strong linear relationship between PCI and AL. Correlation strength between PCI and GE, r = 0.486 (P=0.004), is not as strong as

between PCI and AL. This result shows that increase in PCI is more dependent on increase in AL than increase in GE. The correlation coefficient for GE and LE is, r = 0.541 (p=0.001) which shows moderately strong linear relationship between these variables. It tells us that as LE increase, GE also increases. GE is also correlated with PCI and AL but the strength of this correlation is not very strong. The component plot in Figure 1 also tells us clearly about the correlation between these variables.

Table 6: Bivariate Correlation among HDI indicators

		Zscore: GE	Zscore: LE	Zscore: PCI	Zscore: AL
Zscore: GE	Pearson Correlation	1	.541**	.486**	.402*
	Sig. (2-tailed)		.001	.004	.020
	N	33	33	33	33
Zscore: LE	Pearson Correlation	.541**	1	.230	.211
	Sig. (2-tailed)	.001		.198	.240
	N	33	33	33	33
Zscore: PCI	Pearson Correlation	.486**	.230	1	.811**
	Sig. (2-tailed)	.004	.198		.000
	N	33	33	33	33
Zscore: AL	Pearson Correlation	.402*	.211	.811**	1
	Sig. (2-tailed)	.020	.240	.000	
	N	33	33	33	33

Figure 1: Component plot for HDI indicators



In a second factor analysis we consider another set of indicators, percentage of people not expected to survive to age of 40 years (NETA40), percentage of adults who are illiterate (AI), the percentage of people without access to adequate drinking water facility (PWW) and the percentage of people without access to health services (PWH). These indicators are those occurring in calculation of HPI. These variables were standardized before factor analysis was performed.

Having all these 4 variables in analysis, using the conventional criterion of stopping when the initial eigenvalue drops below 1.0, again 2 of the 4 factors were actually extracted in the analysis (Table 7). These two account for 66% of the variance in the data. We can also see that the first three factors account for 87% of the variance in the data. The extracted two factors have the following form.

Factor 1 = 0.46ZPWH + 0.74ZAI + 0.30ZNETA40 – 0.85ZPWW

Factor 2 = -0.69ZPWH + 0.18ZAI + 0.75ZNETA40 + 0.05ZPWW

Table 7: Extraction of factors

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	1.567	39.178	39.178	1.567	39.178	39.178	1.564	39.098	39.098
2	1.079	26.975	66.154	1.079	26.975	66.154	1.082	27.056	66.154
3	.839	20.986	87.140						
4	.514	12.860	100.000						

The correlation matrix between the variables given in Table 8 tell us that there is one significant inverse correlation between AI and PWW, Pearson's r= -0.420 (P=0.015) which showing moderate inverse relationship between AI and PWW. This relation is difficult to interpret. If we plot these two variables we can see from Figure 2 that actually there are two values causing this inverse relationship. After removing these two variables from our analysis Figure 3 and Table 9 tell us clearly that there is no statistically significant correlation between these two variables. This specific problem

was not observed in the previous analysis. Hence we can conclude that there is no statistically significant correlation exists among AI, PWW, PWH and NETA40.

Table 8: Bivariate Correlation among HPI indicators

		Zscore: PWW	Zscore: PWH	Zscore: AI	Zscore: NETA40
Zscore: PWW	Pearson Correlation	1	-.275	-.420*	-.149
	Sig. (2-tailed)		.121	.015	.407
	N	33	33	33	33
Zscore: PWH	Pearson Correlation	-.275	1	.064	-.070
	Sig. (2-tailed)	.121		.722	.700
	N	33	33	33	33
Zscore: AI	Pearson Correlation	-.420*	.064	1	.105
	Sig. (2-tailed)	.015	.722		.560
	N	33	33	33	33
Zscore: NETA40	Pearson Correlation	-.149	-.070	.105	1
	Sig. (2-tailed)	.407	.700	.560	
	N	33	33	33	33

Figure 2: Scatter plots for AI and PWW before removing two variables

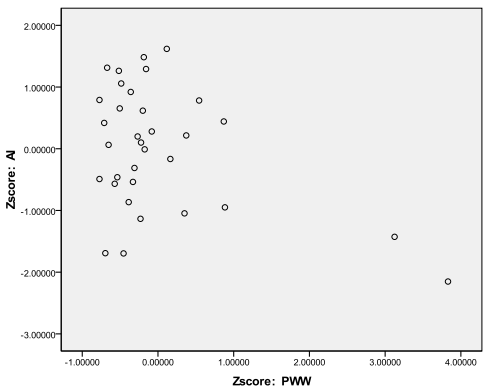


Figure 3: Scatter plot for AI and PWW after removing two variables

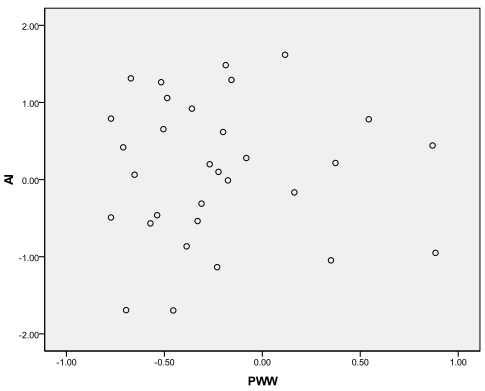


Table 9: Correlation between AI and PWW

		AI	PWW
AI	Pearson Correlation	1	.005
	Sig. (2-tailed)		.980
	N	31	31
PWW	Pearson Correlation	.005	1
	Sig. (2-tailed)	.980	
	N	31	31

4.2 Cluster Analysis

The term cluster analysis first used by Tryon in 1939, encompasses a number of different algorithms and methods for grouping objects of similar kind into respective categories. In many areas, researchers are always interested in finding ways to organize the data in meaningful structure to obtain important information. Cluster analysis is an exploratory data analysis tool which sorts the data into groups in a way that the degree of association between objects is maximal within a group and minimal among the groups. In other words, cluster analysis simply discovers structures in data without explaining why they exist.

A structure of natural grouping is an important exploratory technique for the data. These groups can be used for assessing dimensionality, identifying outliers and suggesting interesting hypothesis concerning relationship.

4.2.1 Cluster analysis methods

Hierarchical and partitional clustering are the two basic methods used for clustering. These two methods are divided into subtypes and different algorithms for finding the clusters. Partitional clustering attempts directly to decompose the data set into a set of different clusters. In partitional clustering a common method for clustering is K-mean clustering. In this method all n observation are partitioned into K cluster in which each observation belong to the cluster with the closest mean. The K-mean method will produce exactly K different clusters of greatest possible distinction. More details about cluster methods can be found in Kaski (1997).

4.2.2 Calculation and Results

In our analysis we carried out K-mean Cluster analysis with all the indicators used for HDI and HPI. However, the percentage of underweight children was excluded as this is a constant value. Also adult illiteracy is excluded since it is related to adult literacy. We observed that the largest values of per capita income and life expectancy are highly affecting cluster sizes. So these two indicators were standardized. After standardizing the two indicators, PCI and LE, the final cluster centers are given in Table 10.

Table 10: Cluster Centers

	Cluster	
	1	2
GE	.9796	.9379
PWW	.1330	.0635
PWH	.1899	.1557
AL	.8203	.6601
Zscore: LE	.62448	-.40591
Zscore: PCI	.90591	-.58884
NETA40	.0759	.1293

The cluster analysis groups all the states of India into two group, first group with 13 states and second group with 20 states. The following groups shows states in these two clusters.

Group 1

Jammu & Kashmir (10,27), Himachal Pradesh (11,4), Chandigarh (1,9), Punjab (9,15) , Sikkim (14,12), Delhi (7,3), Arunchal Pradesh (8,17), Nagaland (3,2), Manipur (22,28), Mizoram (2,14), Goa (4,7), Kerala(6,1), Pondecherry(5,8)

Group 2

Uttaranchal(15,19), Haryana(18,25), Rajasthan(23,33), Uttar Pradesh(32,32), Bihar(31,30), Tripura(20,5), Meghalaya(33,16), Assam(16,6), West Bengal(17,11), Jharkhand(19,20), Orissa(29,23), Chhattisgarh(27,26), Madhya Pradesh(24,29), Gujarat(13,13), Daman and Diu(30,21), Dadra and Nagar Haveli(26,22), Maharashtra(12,10), Andhra Pradesh(28,17), Karnataka(21,24), Tamil Nadu(25,18)

(The figures in the bracket show the HDI and HPI ranks of each state respectively.)

From the above two groups we observe that the 12 states in group 1 are among the 14 top ranks states according to HDI. Only Manipur in this group ranked 22 in HDI. Looking to this cluster we can observe that most of the states in this group are highly developed and rich states of India. Hence it seems that this cluster/group telling us about the rich or developed states of the India.

Table 11: ANOVA for 7 indicators from HDI and HPI

	Cluster		Error			
	Mean Square	df	Mean Square	df	F	Sig.
GE	.014	1	.000	31	40.389	.000
PWW	.038	1	.013	31	2.921	.097
PWH	.009	1	.014	31	.641	.429
AL	.202	1	.009	31	22.455	.000
Zscore: LE	8.365	1	.762	31	10.972	.002
Zscore: PCI	17.603	1	.464	31	37.905	.000
NETA40	.022	1	.007	31	3.212	.083

From ANOVA, Table 11, we can see that PWW, PWH and NETA40 are statistically insignificant. Here it should be notice that these three indicators are used to calculate HPI. After deleting these three variables from our analysis we again perform cluster analysis that give us all the statistically significant variables in ANOVA (Table 12). In this case we obtain two cluster again of same size and have same states as in our first cluster analysis.

Hence we can say that to divide these 33 states into two clusters we only need the indicator from HDI.

Table 12: ANOVA for 4 indicators from HDI

	Cluster		Error			
	Mean Square	Df	Mean Square	df	F	Sig.
GE	.014	1	.000	31	40.389	.000
AL	.202	1	.009	31	22.455	.000
Zscore: LE	8.365	1	.762	31	10.972	.002
Zscore: PCI	17.603	1	.464	31	37.905	.000

5. Discussion and Conclusions

We calculate HDI for all 33 states of India and find that human development is highest in Chandigarh and this state is ranked 1st in HDI and human development is lowest in Meghalaya which ranked 33rd in HDI. We also calculate HPI for all 33 states of India. Kerala and Rajasthan are ranked 1st and last (33rd) in HPI respectively. Multivariate techniques are used to analyze the main indicators used in the calculation of HDI and HPI. In our first analysis we perform factor analysis for the indicators used in the calculation of HDI and find that there are two factors that explain 85% variation in the data. We observe that indicators (PCI and AL) and (GE and LE) are strongly correlated. In our second analysis we perform factor analysis for the indicators used in the calculation of HPI. We observe that three factors explain 87% variation in the data. We did not find any statistically significant correlation between HPI indicators. Using cluster analysis for 7 indicators from HDI and HPI we divide all the 33 states into two clusters. First cluster consist of 13 states and second cluster consist of 20 states. We observe that 12 states in group 1 are among the 14 top ranks states according to HDI. We observe that most of the states are highly developed and rich in 1st cluster. We conclude that 1st cluster represent develop or rich states of India. ANOVA table for these 7 indicators shows that indictors from HPI are statistically not significant. We remove those insignificant variables from our analysis and again perform cluster analysis only for the indicators from HDI. In this case we obtain two clusters again of same size and having same states as in our first cluster analysis. We conclude that to divide these 33 states into two clusters we only need the indicator from HDI.

References

Antony G.M. and Rao K.V. : A composite index to explain variations in poverty, health, nutritional status and standard of living: Use of multivariate statistical methods, Public Health, 2007, vol 121, 578-587

Fitzpatrick J. : Calculating life expectancy and infant mortality rates Mapping health inequalities across London: technical supplement, London Health Observatory 2001.

Johnson R.A. and Wichern D.W. : Applied multivariate statistical analysis, 6th ed. Prentice Hall, 2007

Kaski S. : Clustering methods, Neural Networks Research Centre, Helsinki, University of Technology, EET DST 1997
http://www.cis.hut.fi/~sami/thesis/node9.html [Accessed 27 April 2010]

Haq M.U. : Human Development Report 1990, Oxford university press, New York,1990.

Spearman, C. : 'General intelligence' objectively determined and measured. American Journal of Psychology, 1904, vol 15, 201-293

Tryon, R. C. : Cluster analysis. New York: McGraw-Hill, 1939

United nation development program: http://www.undp.org

Wikipedia: List of states and union territories of India by population

Appendix I
Methodology for calculation of Life Expectancy

Life expectancy is calculated from life table. Life table consist of different columns those are calculated as follows

1. Divide the population in different age groups and note the number of persons in each age group. Then calculate the age specific death rates for each age group by

$$\frac{\text{Number of Deaths in specific age group}}{\text{Nubmer of persons alive in specific age gruop}}$$

2. Calculate width of the each age group interval that is the number of years in each age interval.

3. Calculate average proportion of the year lived by those who die. Generally it is assumed that on an average people will live 0.5 of the interval before death. So for each age group average proportion of the year lived by those who die is 0.5 But in some cases death does not occur uniformly across time within age groups. So for those aged under 1 we assume that the average proportion of the year lived by those who die is 0.1

4. Calculate the probability of dying by the following formula

$$\frac{\text{Number of years in inverval} * \text{age specific death rate}}{1 + \text{number of years in interval} * \text{age specific death rate}}$$

5. Calculate the probability of surviving by subtracting the probability of dying from one. (i.e. 1 – probability of dying)

6. Calculate number of persons alive at the start of the interval. This is a hypothetical population and considers 100,000 alive at age 0. Those alive at next age groups are multiplication of probability of surviving the previous interval and population alive at start of previous.

7. Calculate number of deaths during interval by
Population alive at start of interval – population alive at start of next interval

8. Calculate number of person years lived through the interval by the following

Number of years in interval (number of persons alive at start of next interval + average proportion of year lived by those who die * number of deaths during interval)

Since at age 85+ everybody dies during the interval so an adjustment has to be made. Whatever is used as an estimate of the number of years lived has little impact on overall life expectancy; however, it is usual to use the following estimate:

$$L85+ = \frac{l85}{M85+}$$

9. Calculate total number of person years lived after the interval
This is the 'number of person years lived through the interval' column summed from the bottom.

10. Finally calculate Expectation of life that is the number of years a person aged x can be expected to live

$$\frac{\text{Total number of person years lived after the interval}}{\text{Number of person years alive at the start of the interval}}$$

More detail of the calculation of life expectancy with example is available on website of London Health Observatory.

Appendix II

States and union territories of India by population (As per Census 2001)

Rank	State or union territory	Population	%	Rural Pop.	Urban Pop.	Area km²	Density (per km²)
1	Uttar Pradesh	166197921	16.16	131658339	34539582	240928	690
2	Maharashtra	96878627	9.42	55777647	41100980	307713	315
3	Bihar	82998509	8.07	74316709	8681800	94163	881
4	West Bengal	80176197	7.79	57748946	22427251	88752	903
5	Andhra Pradesh	76210007	7.41	55401067	20808940	275045	277
6	Tamil Nadu	62405679	6.07	34921681	27483998	130058	480
7	Madhya Pradesh	60348023	5.87	44380878	15967145	308245	196
8	Rajasthan	56507188	5.49	43292813	13214375	342239	165
9	Karnataka	52850562	5.14	34889033	17961529	191791	276
10	Gujarat	50671017	4.93	31740767	18930250	196024	258
11	Orissa	36804670	3.58	31287422	5517238	155707	236
12	Kerala	31841374	3.1	23574449	8266925	38863	819
13	Jharkhand	26945829	2.62	20952088	5993741	79714	338
14	Assam	26655528	2.59	23216288	3439240	78438	340
15	Punjab	24358999	2.37	16096488	8262511	50362	484
16	Haryana	21144564	2.06	15029260	6115304	44212	478
17	Chhattisgarh	20833803	2.03	16648056	4185747	135191	154
UT1	Delhi	13850507	1.35	944727	12905780	1483	9340
18	Jammu and Kashmir	10143700	0.99	7627062	2516638	222236	46
19	Uttarakhand	8489349	0.83	6310275	2179074	53483	159
20	Himachal Pradesh	6077900	0.59	5482319	595581	55673	109
21	Tripura	3199203	0.31	2653453	545750	10486	305
22	Meghalaya	2318822	0.23	1864711	454111	22429	103
23	Manipur	2166788	0.21	1590820	575968	22327	97
24	Nagaland	1990036	0.19	1647249	342787	16579	120
25	Goa	1347668	0.13	677091	670577	3702	364
26	Arunachal Pradesh	1097968	0.11	870087	227881	83743	13
UT2	Pondicherry	974345	0.09	325726	648619	479	2034
UT3	Chandigarh	900635	0.09	92120	808515	114	7900
27	Mizoram	888573	0.09	447567	441006	21081	42
28	Sikkim	540851	0.05	480981	59870	7096	76
UT4	Andaman and Nicobar Islands	356152	0.03	239954	116198	8249	43
UT5	Dadra and Nagar Haveli	220490	0.02	170027	50463	491	449
UT6	Daman and Diu	158204	0.02	100856	57348	112	1413
UT7	Lakshadweep	60650	0.01	33683	26967	32	1895
Total	**India**	**1028610328**	**100**	**742490639**	**286119689**	**3287240**	**313**

Appendix III

Frequently Used Acronyms

AL- Adult literacy

ALR-Adult Literacy Ratio

EI- Education Index

GDI –Gross Domestic Income

GDP –Gross Domestic Product

GE-Gross enrolment

GER- Gross Enrollment Ratio

HDI – Human Development Index

HDRs – Human Development Reports

HPI – Human Poverty Index

ICPSR - Inter-university Consortium for Political and Social Research

INCI- Income Index

LE- Life Expectancy

LEI-Life Expectancy Index

NETA40- Percentage of people not expected to survive to age 40

OECD- Organization for Economic Co-operation and Development

PCI- Per Capita Income

PPP –Purchasing Power Parity

PWH- Percentage of people without health facility

PWW- Percentage of people without adequate drinking water facility

UNDP – United Nations Development Program