Identification of Determinant Factors for the Development of C.D. Blocks in Birbhum District: A Multivariate Statistical Approach

Prolay Mondal[a], Ranajit Ghosh[b], Subhasish Sutradhar[c],
[a]Assistant Professor, Department of Geography, Raiganj University, India
[b]Guest Lecturer, Vidyasagar College, Suri, Birbhum, India
[c]Guest Lecturer, Hiralal Bhakat College, Nalhati, Birbhum, India

Abstract

The concept of development is a relative term, but if we want to define the term development it can be defined as improvement of human welfare, quality of life, social well being and satisfying peoples' needs and wants. So a large number of factors in terms of economic, socio-cultural, religion, infrastructural, commercial are involve for the development of a region. In the light of the above aspect an attempt has been made to identify the major factors for the development of Community Development blocks of Birbhum district on the basis of 16 selected variables by multivariate statistical technique, like principal component analysis (method of factor analysis). The study not only aims to identify the major factor of the development but also try to analyze the spatial discrimination of the extracted factors in 19 C.D Blocks of Birbhum District. After analyzing the factor as well as its spatial discrimination, it can be said that all the blocks are not well equipped evenly by four factors. Therefore, the policy makers should focus on each and individual factor separately and find out the exact cause of discrimination.
KEYWORDS: Principal component analysis, Eigen value, factor loading, factor analysis, spatial discrimination.

Introduction:

Geographers become increasingly interested in describing a complex spatial structure of a large number of socio-economic and others variables through some smaller number of underlying dimensions (Aslam Mahmood, 2013). Any administrative area can be well understood with the help of some factorial parameters. In this respect the major factor of development of an administrative area involves the complex interaction between social, cultural, political and economic aspect. The spatial variation in such level of socio-economic development is multidimensional phenomena (Thurstone, 1949, Berry, 1960, Thompson, 1962) can be well interpreted with the help of multivariate statistical approach. Among the different multivariate statistical technique Principal Component Analysis (A branch of factor analysis) is a technique designed primarily to synthesize a large number of variables into a smaller of general components, which retain the maximum amount of descriptive ability. It is a method to discover those hidden factors which might have generated the dependence or covariance among the variables (Morrison, 1967).

In the light of above discussion the present paper is try to bring out the major factors for the development of different C.D. Blocks of Birbhum District as well as to analyze the spatial variation of such factors or components of the respective district.

2. Materials and Methods:

2.1 Study Area:

Birbhum District is located between $23^0$32'30"N to $24^0$35'00"N and $87^0$5'25"E to $88^0$01'40"E (Fig.1). This district lies at the easternmost part of the West Bengal at the border of Jharkhand (Fig.1). There are 19 Community Development (C.D) Blocks with 3502387 population (2011) among which 12.80% reside in urban area and 87.2% reside in rural area.

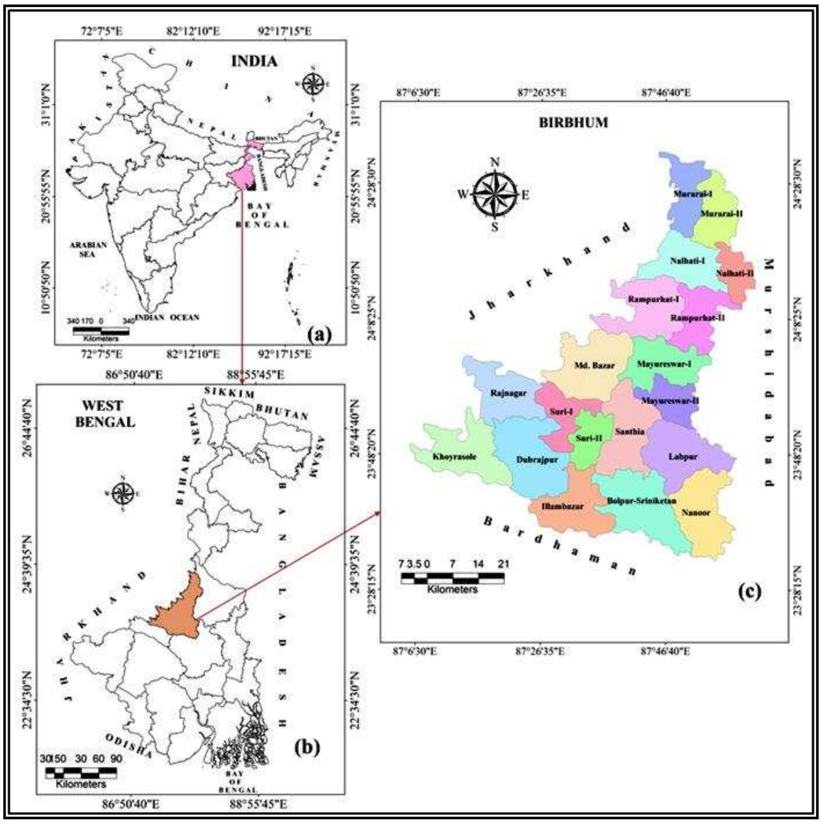


Fig.1

(a) India, (b) West Bengal, (c) Birbhum

2.2 Objectives:

The main objectives of the study are----

- To find out the main factors behind the development of the district.
- To evaluate the spatial variation of those factors in different C.D. Blocks of Birbhum District.

2.3 Data base and Methodology:

The whole work has been done with the help of secondary data which are collected from District Statistical Handbook (2010-2011), Statistical Handbook West Bengal (2010) and District Census Handbook (2001).

In order to identify the major factors of development as well as their spatial variation in C.D. Blocks of Birbhum District, Principal Component Analysis as a method of Factor

Analysis has been used here. The main aim of Principal Component method is the construction out of given set of variables X_j's (J=1,2….K) of new variables (p_1), called principal component which are linear combination of X's (Hotelling, 1933).

Selection of variables:

Table 1: Selection of Variables

Sl No.	Denotation	Variables
1	X_1	Cultivated area
2	X_2	Total irrigated area
3	X_3	Seed stores
4	X_4	Fertilizer depots
5	X_5	Electrified mouzas
6	X_6	Mouzas having drinking water facilities
7	X_7	Length of the road
8	X_8	Cultivators
9	X_9	Agricultural laborers
10	X_{10}	Household industrial workers
11	X_{11}	Banks
12	X_{12}	Medical institution
13	X_{13}	Doctors
14	X_{14}	Literacy rate
15	X_{15}	Primary School
16	X_{16}	High School

To extract the major components or the major factors, different variables from different sectors like agriculture, education, employment, transport, banking, amenities, facilities and health have been selected. The following list of 16 variables together with their denotation is provided in the Table 1.

3. Results and Discussion:

Principal Analysis starts with the correlation matrix (Table 2) of original data set of sixteen variables because the original data set is not in standardized format. So to standardize the data set correlation matrix has been done as an input of PCA as well as to extract the factors.

Table 2: Correlation Matrix

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}
x_1	1.000															
x_2	.243	1.000														
x_3	.214	.482	1.000													
x_4	-.048	.559	.658	1.000												
x_5	.784	.408	.407	.150	1.000											
x_6	.793	.465	.463	.191	.981	1.000										
x_7	.451	.100	.061	-.099	.457	.459	1.000									
x_8	.583	.482	.657	.491	.638	.649	.143	1.000								
x_9	.708	.472	.537	.264	.546	.599	.353	.809	1.000							
x_{10}	-.097	-.185	-.388	-.233	-.517	-.490	-.296	-.260	-.033	1.000						
x_{11}	.396	.138	.446	-.015	.383	.447	.229	.207	.425	-.250	1.000					
x_{12}	.647	.226	.350	-.097	.496	.506	.473	.465	.674	-.161	.692	1.000				
x_{13}	.435	.147	.582	.022	.277	.292	.175	.432	.551	-.165	.586	.806	1.000			
x_{14}	-.083	.448	.545	.674	.252	.262	-.119	.355	.139	-.605	.258	-.019	.036	1.000		
x_{15}	.775	.347	.594	.334	.789	.802	.438	.830	.830	-.276	.431	.564	.447	.250	1.000	
x_{16}	.598	.348	.264	.299	.455	.460	.303	.717	.775	-.016	.328	.500	.237	.204	.726	1.000

Source: Calculated by the authors

After standardization of the data the next task is to find out major components which account for a large proportion of the total variance explained. There are some rules for determining how many factors should be retained (Field 2000, Rietveld and Van Hout 1993) for analysis.

Rule 1: Retain only those factors which have the Eigen value greater than 1 (Guttman-Kaiser rule).

Rule 2: Keep only those factors which account for about 80% of the total variance.

Table 3: Total Variance Explained

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	7.421	46.381	46.381
2	2.574	16.087	62.468
3	1.517	9.482	71.950
4	1.502	9.385	81.334
5	.698	4.364	85.698

6	.651	4.070	89.768
7	.619	3.870	93.637
8	.402	2.513	96.150
9	.205	1.280	97.430
10	.145	.904	98.335
11	.096	.598	98.933
12	.070	.440	99.373
13	.051	.316	99.689
14	.027	.171	99.860
15	.017	.108	99.968
16	.005	.032	100.000

Source: Calculated by the authors

To get a realistic result, the components, eigen value and their percentage of variance have been calculated in the Table 3. According to the above rules it can be interpreted that only the first four component are extracted (Table 3) because it explain 81.334% (the first principal components accounts for 46.38% of the total variance; second accounts for 16.08% and third and fourth further accounts for 9.48% and 9.38% of the total variance) of the total components. Rests of other components have been eliminated. Therefore, the first four factors are most heavily concentrated and those are our concern for result analysis.

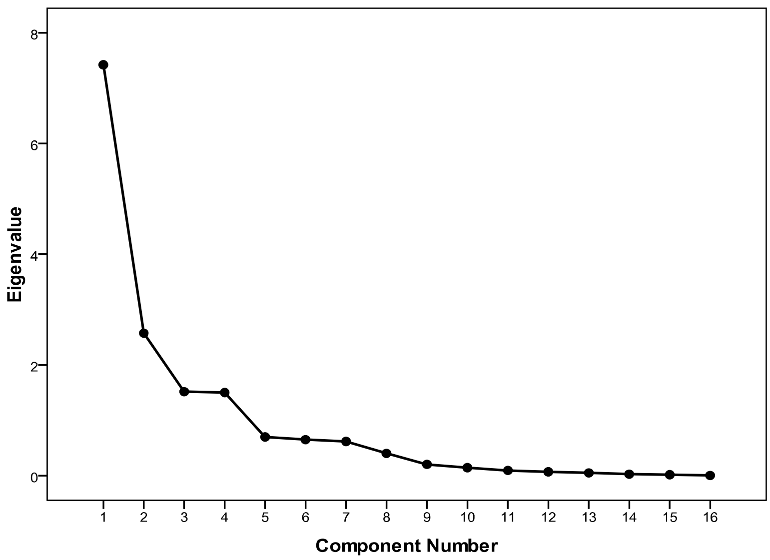


Fig. 2 Scree Plot for Eigen value of Factors

The graphical presentation of eigen value in respect of factors is called Cattle's (1966) Scree Test (Fig. 1). It is a simple line segment plot that shows the fraction of total variance in the data. It shows that when there are substantive factors, the slope of the line will be steep but when the factors correspond to the error, the slope will be flat. In fig1 first four factors are substantive factors as it formed steep slope, and rest of other is error factors as it formed flat slope.

In principal component analysis, the variables are rotated to obtain new variable. Rotation is a means for estimating factor loading so that factors are easily interpretable. Varimax rotation is mostly used in principal component analysis, it produces uncorrelated factors. The result of PCA after varimax rotation is given below.

Table 4: Varimax Rotated Factor Loading

Variable	Factor1	Factor 2	Factor 3	Factor 4
x_1	.675*	-.175	.552*	.276
x_2	.436	.572*	.109	-.008
x_3	.281	.723*	.055	.485
x_4	.297	.851*	-.140	-.118
x_5	.440	.209	.812*	.144
x_6	.466	.238	.780*	.179
x_7	.197	-.214	.661*	.151
x_8	.767*	.435	.203	.191
x_9	.840*	.139	.161	.398
x_{10}	.290	-.543*	-.678*	-.155
x_{11}	.071	.115	.262	.799*
x_{12}	.403	-.119	.300	.790*
x_{13}	.233	.067	.017	.904*
x_{14}	-.059	.892*	.143	.052
x_{15}	.729*	.239	.466	.284
x_{16}	.823*	.111	.148	.130

Rotation Method: Varimax with Kaiser Normalization.

* indicate highly positive and negative value in each factor and these are taken as major factors

Source: Calculated by the authors

Now the varimax rotated factors are helped to identify the major factors for the development of Birbhum district. Here only the strong positive and strong negative correlation value have been taken into consideration from each factor for our analysis. According to Stevens (1992) “recommends interpreting only factor loading with absolute value greater than 0.4 (which explain around 16% of variance)”. So here values of 0.5 and more than 0.5of each component are chosen for analysis.

Factor-1: The factor loading in the first principal component show that it has significant positive correlation with cultivated area in hectares **(X_1)**, numbers of cultivators

(X_8),number of agricultural laborers **(X_9).** As this variable is the consequences of agriculture therefore it can be call agricultural development. On the other hand in the first factor the another correlation can be seen in number of Primary School **(X_{15})** and number of high school **(X_{16})** which is the consequences of Educational Infrastructure Development. So the first factor for the development of the C.D. Blocks of Birbhum district is combindly identified as Agriculture **and Educational Infrastructural Factor.**

Factor-2: In the second component three values load heavily which area irrigated are **(X_2)**, number of seed stores **(X_3)**, number of fertilizer depots **(X_4)**, literacy rate **(X_{14})**. It indicates that with increasing of literacy the famers are conscious about the use of modern output of agriculture. On the other hand, in this component number of household industrial workers **(X_{10})** shows negative value this means the application of modern output in agriculture are the result of high agricultural output, as a consequence the household industrial workers gradually decrease and they have tend to participate in agricultural sector. Therefore the second factor is used to call
Modernization in Agricultural Sector.

Factor-3: This factor includes three variables with high positive value and one variable with high negative value i.e. number of mouzas electrified**(X_5)**, number of mouzas having drinking water facilities **(X_6)**, length of the road in Km **(X_7)** have high positive value and number of household industrial workers **(X_{10})** has high negative value. This indicates with the development of infrastructure there is shifting of household industrial workers to the other sectors. So, this factor can be called **Level of infrastructural Development.**

Factor-4: In the fourth principal component three factors are heavily concentrated which are number of doctors **(X_{13})**, number of medical institution **(X_{12})** and number of banks **(X_{11}).** This factor indicates **Health and Commercial Development**.

Spatial Discrimination of factors in different C.D. Blocks: It is also an important concern to identify the spatial discrimination of factors in different C.D Blocks of Birbhum district because all the factors are not evenly concentrated throughout the blocks.

Spatial Discrimination of factor-1: The first factor is Agricultural and Educational Infrastructural Factor. Fig.3 shows the concentration of agriculture and educational infrastructure is higher in Bolpur-Sriniketan, Sainthia, Labpur, Durajpur, Nanoor, Mayureswar-I, Rampurhat-II, Nalhati-I and Murarai-II blocks. All the above blocks have large amount of cultivated area as well as cultivators & numbers of school due to their larger area (ranges between 176.78 Km^2 to 342.21 Km^2). But the blocks like Rajnagar, Suri-I and Suri-II have low level of agriculture and educational infrastructure (Fig.3) In case of Suri-I and II though it is a develop block but the area is too small (154.83 Km^2 and 135.85 Km^2 respectively) to hold large number of cultivated area and educational infrastructure in respect of other blocks. Rajnagar occupy its low level character due to its unfavorable geomorphic and climatic situation (part of Ranchi plateau and drought prone block). The rest of blocks have moderate level of agriculture and educational infrastructure.

Spatial Discrimination of factor-2: This factor is named Modernization in Agricultural Sector. This is more or less depending upon the first factor. The blocks which have higher level agricultural advantages are involved for its modernization as well as the educational infrastructure (Factor-1) influences the literacy which ultimately affect in agricultural modernization. On the basis of these Illambazar, Labpur, Sainthia, Mayureswar-I and II, Rampurhat-I and II, Nalhati-I and II have the higher level of agricultural modernization (Fig.4). According to result though Suri-II has low agricultural advantages (Fig.3) but due to high literacy it achieves higher level of agricultural modernization (Fig.4). On the other hand though Bolpur-sriniketan block have higher level of agriculture advantages (Fig.3) and educational infrastructure but it has low level of modernization in agricultural sector (Fig.4) due to poor implementation of modern agricultural input.

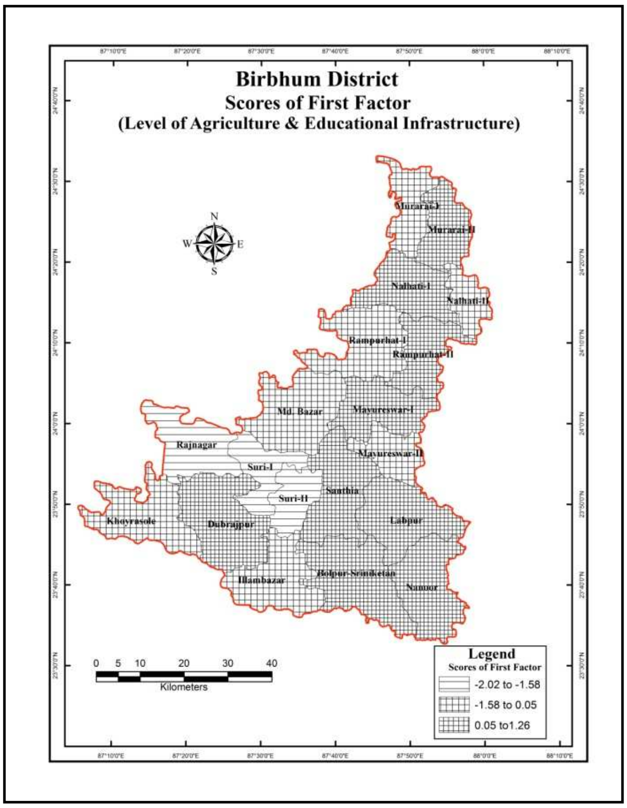


Fig.3

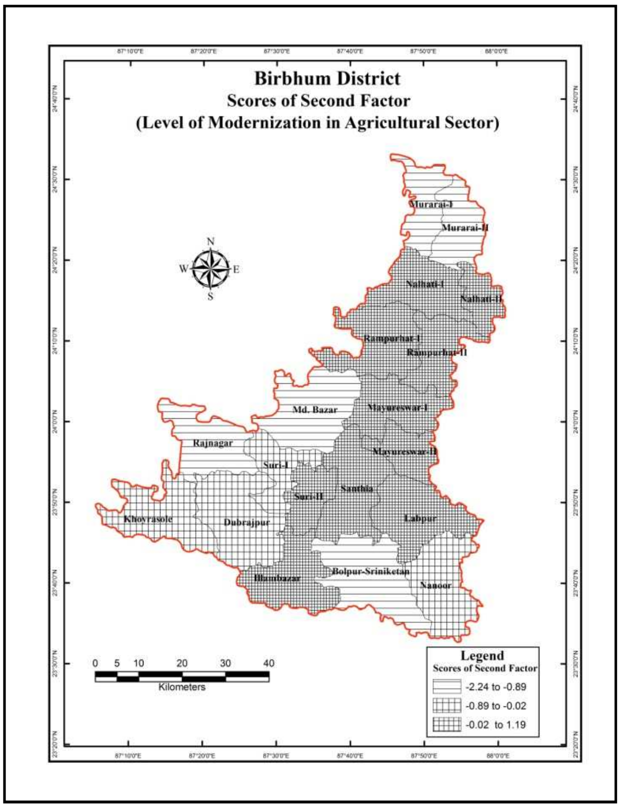
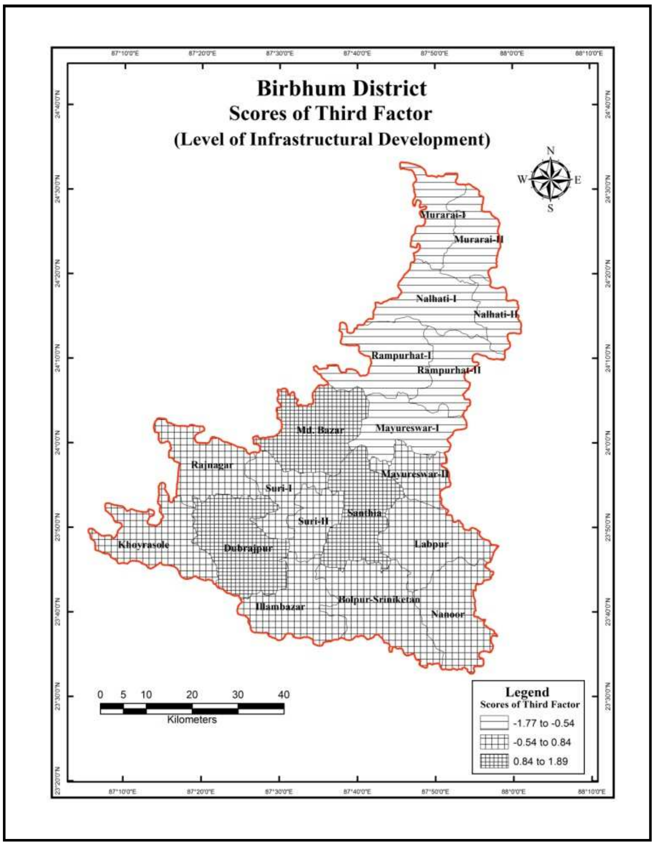


Fig.4

Spatial Discrimination of factor-3: This factor identified as Level of infrastructural Development. Fig.4 shows that only three blocks i.e. Mohammad Bazar, Sainthia & Dubrajpur provide higher level infrastructure and the blocks like Murarai-I & II, Nalhati-I & II, Rampurhat-I & II and Mayureswar-I have low level of infrastructure. Rests of the blocks have moderate level of infrastructure (Fig.5).

Spatial Discrimination of factor-4: This factor is treated as Health and Commercial Development. In this factor also acute discrimination is found. Only the blocks like

Bolpur-Sriniketan, Rampurhat-I as well as Suri-I and Dubrajpur have higher to moderately higher level of health and commercial development (Fig.6) because in Blopur, Rampurhat, Suri and Dubrajpur have more or less well facilated government hospital and nursing home due to its urban character and also have different government and non-government bank branches. But the rest of the blocks have low level of health and commercial development (Fig.6). Regarding the health, lower develop blocks are depended on the their nearest block having government hospital and nursing home e.g. Murarai-I & II, Nalhati-I &II and Rampurhat-I depend on the hospitals and nursing home of Rampurhat-I block.



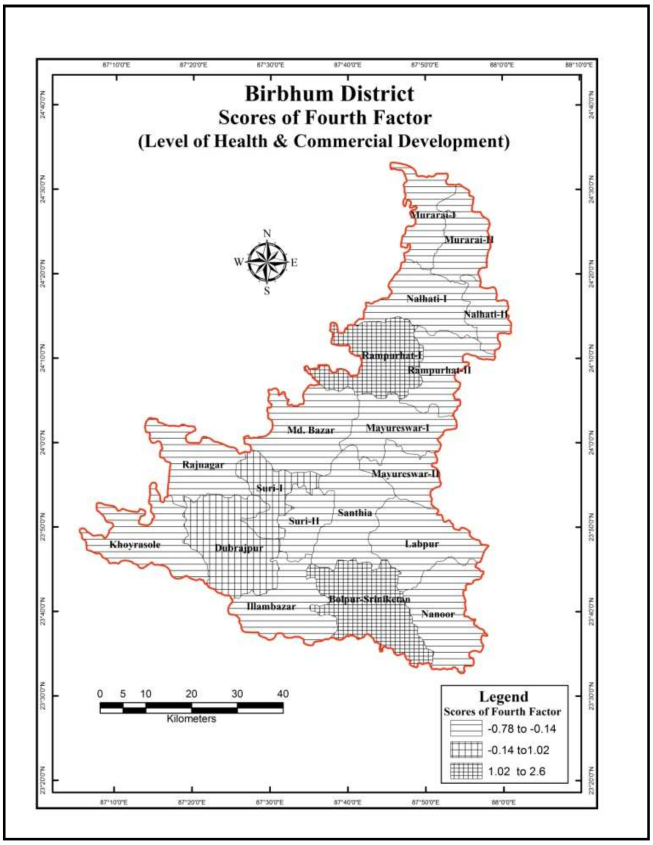


Fig. 5 **Fig.6**

4. Conclusion: This paper reveals empirical result of factor analysis to identify the major factor for the development of C.D Blocks of Birbhum District. This analysis carried out with limited number of variables but can be extended to a large number of variables. After analyzing the factor as well as its spatial discrimination, it can be said that all the blocks are not well equipped evenly by four factors. Some blocks are well facilated by first factor or some other by second factor and so on. Therefore, the policy makers should focus each and individual factor separately and find out the exact cause of discrimination of each factors and put their effort particularly on the lagged blocks, so that all the blocks and ultimately Birbhum district will be reached at the highest peak of the development.

5. Acknowledgements: Authors express gratitude to Dr. Gopal Chandra Debnath, Former Professor, Department of Geography, Visva-Bharati, Santiniketan for his constant supervision, invaluable keen interest, wise counseling and encouragement. We also thankful to Dr. Niladri Das for his technical cooperation to complete this research work.

References:

Anderson, T.W. (1958) Introduction to Multivariate Statistical Methods, John Wiley & Sons, New York.

Berry, B.J.L. (1960) An Inductive Approach to The Regionalization of Economic Development, in Ginsburg, N. (ed.), Essays on Geography and Economic Development, Chicago University, pp 78-100.

Cattle, R.B. (1966) The Scree Test for the Number Factors, Multivariate Behavioral Research, pp 245-276.

Cooley, W.W. and Lohnes, P.R. (1971) Multivariate Data Analysis, John Wiley and Sons, New York.

District Census Handbook, Birbhum District, 2001.

District Statistical Handbook, Birbhum District, 2010-2011.

Field, A. (2000) Discovering Statistics Using SPSS for Window, London-Thousand Oaks-New Delhi: Sage Publication.

Hotelling, H. (1933) Analysis of a Comples Statistical Variables into Principal Component, Journal of Educational Psychology, Vol.24, pp-407-441.

Kaiser, H.F. (1958) The Vaimax Criterion for Analytic Rotation in Factor Analysis, Psychometrica. Vol.23, No.3. 1958.

Mahmood Aslam (2013) Statistical Method in Geographical Studies, Rajesh Publication, New Delhi, pp-159-168.

Morrison, D.F. (1967) Multivariate Statistical Method, McGraw-Hills Inc., New York, pp-221-225.

Riveld, T and Van Hout, R. (1933) Statistical technique for the Study of Language and language Behaviour. Belin-New York: Mouton de Gruyter.

Thompson, J.H. (1962) Towards A Geography of Economic Health: The Case of New York State, Annals of Associationof American Geographer, pp-1-20.

Thurstone, L.L. (1949) Multi Factor Analysis. University of Chicago Press, Chicago.