



Mapping Demographic Diversity in Indian States: an Analysis of 2011 Census

R.GEETHA

ASSOCIATE PROFESSOR DEPARTMENT OF STATISTICS
S.D.N.B.VAISHNAV COLLEGE FOR WOMEN CHROMEPET CHENNAI-44

S.SANTHANA
LAKSHMI

RESEARCH SCHOLAR DEPARTMENT OF STATISTICS S.D.N.B.VAISHNAV
COLLEGE FOR WOMEN CHROMEPET CHENNAI-44

ABSTRACT

This paper proposes to explore whether natural grouping or cluster exists with respect to the distribution of child (0-6yrs), adolescent (15-24yrs), adults (25-49yrs), and old age (above 65 yrs.) population groups across the states in India using census data 2011. Partitioning Around Medoids is used to map the diversity among the states in India by identifying homogeneous subgroups. Socio - demographic profiles of the clusters formed are highlighted. Child Sex ratio is found to be statistically significant ($p < .05$) across clusters using ANOVA

KEYWORDS : Partitioning around medoids, census, ANOVA

1. INTRODUCTION

India in last 50 years has huge growth in population and has become most populated when compared with other countries all over the world. According to Census reports of India 2011 the total population of India has reached to 121.6 crores with 62.31 crores males and 58.74 crores female. The density of population per sq. km. is 382 persons. The Sex ratio is 940 females per 1000 males and the child sex ratio is 914 females per 1000 males. The Literacy rate of India has gone up to 73.0 percent from previous figure of 64.83 percent. The rural population constitutes 68.8% and urban population 31.2% o. The proportion of rural population declined from 72.19% in 2001 Census to 68.84% in 2011 Census. Phase of urbanization has been steadily growing in India from 27.81 % in 2001 Census to 31.16 % in 2011. The phenomenal increase in census in towns was also responsible for the increase in urbanization.

Adolescence is a crucial period of development in terms of education and professional career in various fields of socio-economic and demographic indicators. Due to the demographic dividend among Indian states, there is a considerable difference in the distribution of child and adolescent population. As per Census 2011, the child population in the age group of 0-6 years stands at 16.45 crores and adolescent population at 23 crores. It is of interest to map the demographic dividend across states by bringing together states with similar characteristics.

Previous studies have reported clustering of socio economic behaviors, human risk behaviors to determine if cluster differed at two different time points, by gender and by socio-economic status, M.VishnuVardhana Rao, et al. (2013) in a study undertaken in Uttar Pradesh, explained the concepts of variations between clusters in their socio-demographic and health nutrition factors. Aneta Piekut et al. (2012) describes multi-dimensional social diversity in the two European cities namely, Leeds and Warsaw based on the demographic and socio economic factors , using K-means algorithm. Wiwandari Handayani (2013) uses the multi-variate cluster analysis to understand the rural-urban transition in central Java.

This paper aims to identify and capture the spatial variation among Indian states with respect to the age-wise distribution of population and child sex ratio. The significant demographic factors between the clusters formed are highlighted.

2. DATA AND METHODOLOGY

Census is the process of collecting, compiling, analyzing and distributing demographic, social, cultural and economic data relating to all persons in the country, in the period of ten years interval. In this study, data for 35 states in India from census 2011 are considered for the variables presented in Table 1.

Table 1: Description of variables

Variables	Description
Child population	Proportion of children (0-6)years
Adolescent population	Proportion of adolescents (15-24)years 15-2415-24population
Adult population	Proportion of adults (25-49)years
Old age population	Proportion of old age people (65 and above)
Child Sex Ratio	Girls per thousand boys
Urban population	Proportion of people living in urban areas

2.1 Cluster analysis:

Cluster analysis is designed to generate subgroups from a sample of respondents that represent genuine within cluster homogeneity while maximizing between cluster differences. Hierarchical clustering and K-means clustering by Mac Queen are the popular methods of clustering in literature.

Partition Around Medoids (Kaufman and Rousseeuw, 1987) is a robust method of clustering that uses real objects (medoids) to represent the cluster. The algorithm is Select k representative objects arbitrarily

For each pair of non-selected object h and selected object i , calculate the total swapping cost $TC_{ih} = d(j,h) - d(j,i)$ where $d(j,i)$ is a distance measure of similarity

For each pair of i and h ,

If $TC_{ih} < 0$, i is replaced by h

Then assign each non-selected object to the most similar representative object repeat steps 2-3 until there is no change

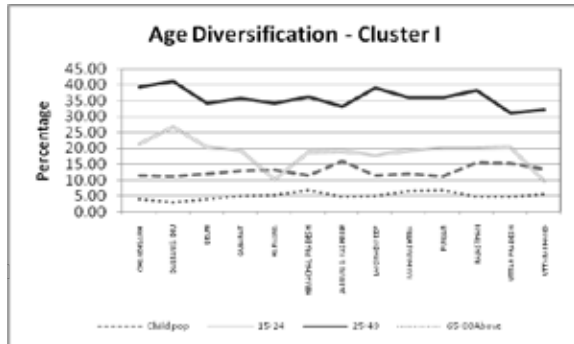
3. RESULTS:

The variables in Table 1 are used for clustering Indian states using Partition Around Medoids algorithm. Two distinct clusters of sizes 13 and 22 are identified. Cluster 1 contains 13 states - Chandigarh, Daman & Diu, Delhi, Gujarat, Haryana, Himachal Pradesh, Jammu and Kashmir, Lakshadweep, Maharashtra, Punjab, Rajasthan, Uttar Pradesh, and Uttarakhand. Cluster II comprises of 22 states - Andaman & Nicobar Island, Andhra Pradesh, Arunachal Pradesh, Assam, Bihar, Chhattisgarh, Dadra Nagar Haveli, Goa, Jharkhand, Karnataka, Kerala, Madhya Pradesh, Manipur, Meghalaya, Mizoram, Nagaland, Odisha, Puducherry, Sikkim, Tamil Nadu, Tripura, and West Bengal. A clear distinction exists in the clustering pattern as evident from the cluster centroids given in Table 2.

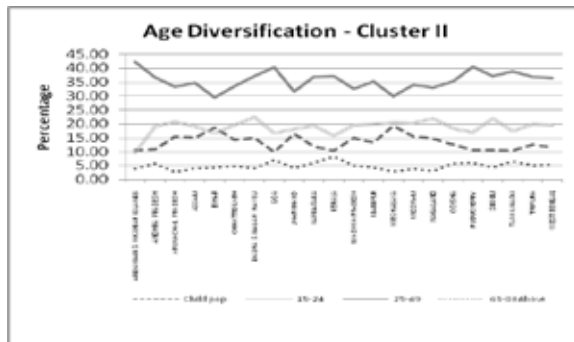
Table 2: Value of Cluster Centroids

Variables	Cluster	
	1	2
	size=13	size=21
Age 15 – 24	18.69	18.84
Age 25 – 49	35.84	35.58
Age 65 above	5.13	5.00
Childpop	12.82	13.32
Urban population	47.93	33.63
Child Sexratio	883	951

The cluster centroid for child population in cluster 2 is 13.32 which is higher than that of cluster 1 whose centroid is 12.82. Wide disparity in child sex ratio exists between the clusters. Cluster 2 has maximum of 951 girls per thousand boys compared to cluster 1 which has 883 girls per thousand boys. But cluster 1 has 47.93 as urban centroid which is higher than that of cluster 2. Urbanization is more prominent in cluster 1. The distribution of adolescents and adults are almost same in the two clusters while old age distribution is more in cluster 1 than in cluster 2.

Fig.1 and Fig.2 shows the age distribution of population across Clusters.

- Jammu & Kashmir has the highest percentage in the age group 0-6 years of 16.10 percent whereas Punjab and Daman & Diu has the lowest percentage age group of 11.10.
- Daman and Diu has the highest percentage age group 15-24 of 26.68 percent whereas Uttarakhand has the lowest percentage age group of 9.72 percent.
- Daman and Diu has the highest percentage age group 25-49 of 41.1 percent whereas Uttar Pradesh has the lowest percentage of 31.05 percent.
- Himachal Pradesh has the highest percentage in age group 65 above of 6.87 percent whereas Daman and Diu has the lowest percentage of 2.9 percent.

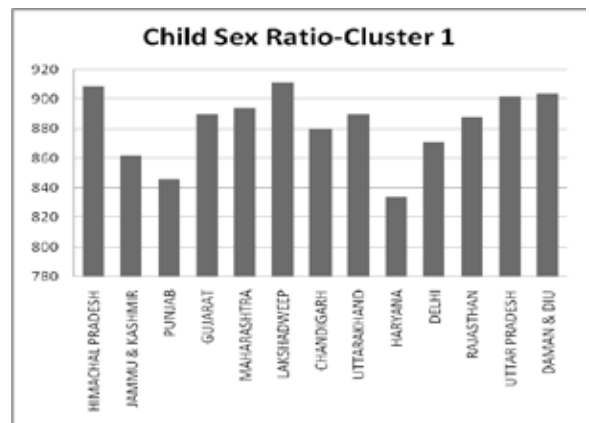
Fig.2

- Meghalaya has the highest percentage age group 0-6 is 19.2 percent whereas Goa has the lowest percent 9.9.
- Dadra & Nagar Haveli has the highest percentage age group 15-24 is 22.52 percent whereas Andaman & Nicobar has the lowest percent 9.75.
- Andaman & Nicobar Island has the highest percentage age group 25-49 of 42.27 percent whereas Bihar has the lowest percent 29.41.

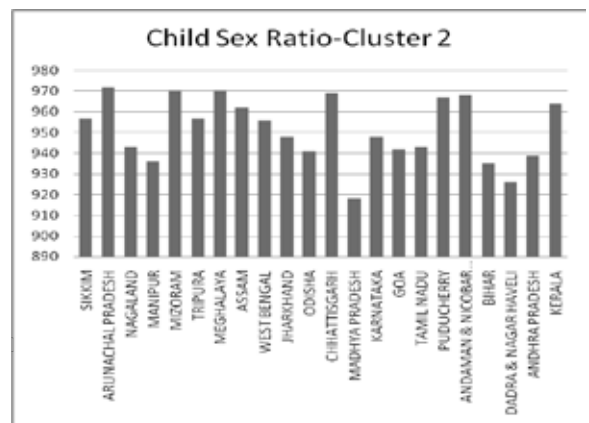
- Kerala has the highest percentage age group 65 and above of 8.32 percent whereas Arunachal Pradesh has the lowest percent 2.76.

3.1 Child Sex ratio

A comparison of the child sex ratio across clusters formed is shown in Fig 3 and Fig.4

Fig.3

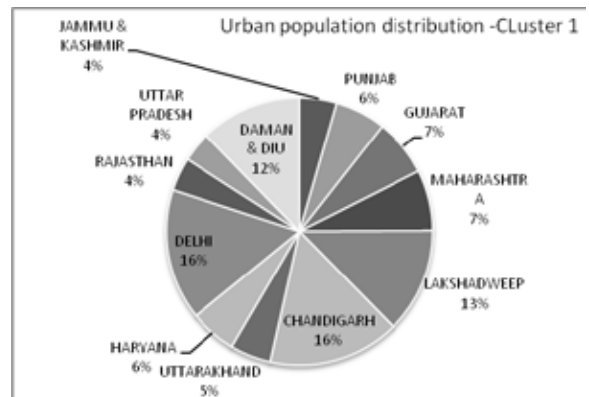
Lakshadweep has the highest Child sex ratio 911 and Haryana has the lowest Child sex ratio 834.

Fig.4

Arunachal Pradesh has the highest Child sex ratio 972 and Madhya Pradesh has the lowest Child sex ratio of 918.

3.2 Urban population distribution

A comparison of the distribution of urban population among the clusters formed is shown in Fig 5 and Fig.6.

Fig.5

Delhi has the highest urban population percentage 97.5 and Himachal Pradesh has the lowest population percentage of 10.0.

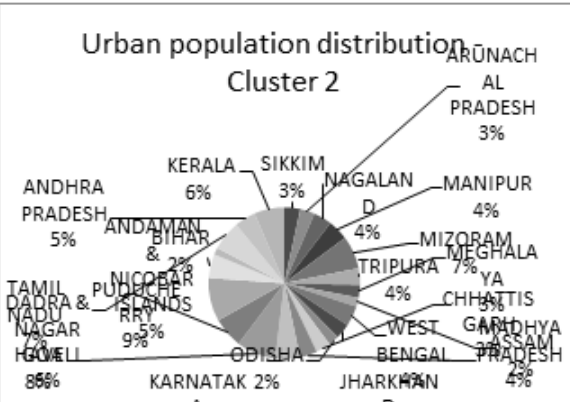


Fig.6

TABLE 2: Test for equality of age-wise population means

Variables		Sum of Squares	df	Mean Square	F	Sig.
Age15-24	Between Groups	.189	1	.189	.016	.900
	Within Groups	390.337	33	11.828		
	Total	390.526	34			
Age25-49	Between Groups	.557	1	.557	.054	.817
	Within Groups	338.817	33	10.267		
	Total	339.375	34			
Age65 and above	Between Groups	.144	1	.144	.086	.771
	Within Groups	55.264	33	1.675		
	Total	55.408	34			
Child population	Between Groups	2.066	1	2.066	.355	.555
	Within Groups	191.790	33	5.812		
	Total	193.855	34			
Urban	Between Groups	1671.793	1	1671.793	3.661	.064
	Within Groups	15069.011	33	456.637		
	Total	16740.804	34			
childsexratio	Between Groups	38068.875	1	38068.875	105.119	.000
	Within Groups	11951.010	33	362.152		
	Total	50019.886	34			

4 Summary and Conclusion:

PAM method of clustering states in India results in two diversified clusters of size 13 and 22.

The characteristic of cluster 1 is that the states are more urban with lower proportion of Child in (0-6 yrs.) and more dependent population with lower child sex ratio.Cluster 2 comprises of states with less proportion of dependents, higher proportion of child population and low concentration in urban areas. Child sex ratio is higher in cluster 2.Among the states in cluster 1 Daman and Diu with population density 2191 per sq.km has highest 41.1% adolescent population and lowest percentage of old age population In cluster 2 Andaman and Nicobar island with a population density of 46 per sq.km has higher adolescent population and Arunachal Pradesh with density 17 per

Puducherry has the highest urban population 68.3% and Bihar has the lowest population 11.30%It clearly shows that Urbanization is more prominent among cluster 1 states .

Analysis of Variance is used to find if significant difference exists in the age-wise population groups between clusters. The results are presented in Table 2

Child Sex ratio is found to be statistically significant (p<0.05).The distribution of girls per thousand boys is more in cluster 2 than in cluster 1 states. This information helps in effective planning of policies for girl children to make them responsible citizens.

sq.km. has the lowest percentage of old age population

The highlights of the study is that the child sex ratio is highly statistically significant (p<0.001) between the clusters. This study brings about the spatial variation in Indian states with respect to age-wise population distribution. Such a mapping of the demographic diversity will help in studying the clusters separately and to frame policy and enables quick decision making.

REFERENCES

1. Aneta Piekut, Philip Rees et al. (2012) Multidimensional diversity in two European cities:thinking beyond ethnicity Environment and Planning A , volume 44, pages 2988 – 3009 | 2. Wiwandari Handayani, Rural-Urban transition in Central Java: Population and Economic Structural changes Based on Cluster Analysis, Land 2013,2,419-436 | 3. J.F.Hair, W.C.Black et.al. , Multivariate Data Analysis, Pearson, Sixth edition,2011 | 4. M.Vishnu-Vardhana Rao et.al. ,A study of geographical clustering of districts in Uttar Pradesh using nutritional anthropometric data on pre-school children, Indian J Med Res 2013,pp 73-81 | 5. www.censusindia.gov.in. |