

Ranking of Districts of West Bengal by an Index Developed Through Census 2011 Household Data (Urban & Total)

Partho Pratim Bhadra¹, Tushar Kanti Ghara²

¹Deputy Director, Directorate of Census Operations, West Bengal, Govt of India
IB-199, Sector –III, Salt Lake, Kolkata 700106

²Joint Director of Public Instruction, Department of Higher Education, Govt of West Bengal
8th Floor, BikashBhavan, Salt Lake City, Kolkata 700091

Corresponding author: Tushar Kanti Ghara

ABSTRACT: Housing condition, Amenities and Assets are the three broad categories of Census Data for House listing and Housing Census. Here, an attempt has been made to use two different approach for finding Ranks of districts based on these House listing and Housing Census Data for Urban West Bengal and separately for Total dataset. The districts were ranked using two different approach and statistically tested (Bhadra & Ghara). Indices have been developed for districts and accordingly the districts ranked with the data for Urban and Total. Statistical techniques like Principal Component Analysis, and Wilcoxon Signed Ranks Test, Sign Test and other Descriptive Statistical theories have been applied on this. Since this study is for Urban and Total, all the 19 districts as per Census 2011 are included.

KEYWORDS: Principal Component Analysis (PCA), Rotated component matrix, Sign Test, Wilcoxon Signed Ranks Test

Date of Submission: 09-07-2020

Date of Acceptance: 25-07-2020

I. INTRODUCTION

The Census of India 2011 was the 15th modern Census in India. It was the seventh after Independence. Census in India is conducted in two phases, (i) House listing & Housing Census is the 1st Phase and (ii) Population Census is the 2nd phase. 1st Phase helps to prepare the frame for collection of information during next phase. An attempt is being made to rank the districts of the state of West Bengal on the basis of Housing infrastructure, Availability of Household Amenities and Assets. The ranking of the districts have been calculated based on the individual values of each observations. Ranks obtained by adopting the two approaches have been tested statistically to check the significance of the rank. **The data analysis has been done using STATA and SPSS.**

The urban area for Census of India 2011 consists of two categories. The first one is known as Statutory Towns and the second one is known as Census Towns. Statutory towns are basically notified town area under law by concerned State which may be Corporations, Municipality or Cantonment board etc. Two main characteristics of statutory towns are Outgrowth (OG) and Urban Agglomeration (UA). OG is a unit like village or hamlet which is clearly identifiable in terms of its boundaries and location. Military camps, railway colonies, University campus, Port area etc are some examples of OG if these areas come up near a Statutory Towns outside its statutory limits but within the limit of a revenue village /villages contiguous to the town. Presence of urban features in terms of infrastructure and amenities such as pucca roads, electricity, taps, drainage system for disposal of waste water etc. educational institutions, post offices, medical facilities, banks etc. is ensured while determining the OG of a town. The second characteristic is Urban Agglomeration (UA). An urban agglomeration is a continuous urban spread constituting a town and its adjoining outgrowths (OGs), or two or more physically contiguous towns together with or without outgrowths of such towns. An Urban Agglomeration must consist of at least a statutory town and its total population (i.e. all the constituents put together) should not be less than 20,000 as per the 2001 Census. In varying local conditions, there were similar other combinations which have been treated as urban agglomerations satisfying the basic condition of contiguity (Census 2011). Places that satisfy the following criteria are termed as Census Towns (CT). (a) minimum population of 5000 (b) At least 75% of the male main working population engaged in non-agricultural pursuits (c) A density of population of at least 400 per sq.km.(Census 2011).

II. DATA

Census of India 2011 is used for the purpose of the study. The exercise is done on the Urban data set and combined data of Rural & Urban, i.e., Total Dataset of House listing & Housing Census which was conducted in 2010. The main variables are Housing infrastructure, Availability of Household Amenities and Assets. We consider here 11 such main variables and 79 sub variables. These 79 variables are actually sub variables of 11 main variables. The variables are – Uses of Census house (residence, residence-cum-other use, shop/office, school/ college/etc, hotel/ lodge/guest house/etc, hospital/ dispensary/etc, factory/workshop/ work shed/etc., place of worship, other non-residential use); condition of census houses (good, livable, dilapidated); material of roof (grass/thatch/bamboo/wood/mud/etc, plastic/polythene, handmade tiles, machine made tiles, burnt brick, stone/slate, GI/metal/asbestos sheets, concrete, any other material); material of wall (grass/thatch/bamboo/wood/mud/etc, plastic/polythene, handmade tiles, machine made tiles, burnt brick, stone/slate, GI/metal/asbestos sheets, concrete, any other material); material of floor (mud, wood/bamboo, burnt brick, stone, cement, mosaic/floor tiles, any other material); availability of assets (radio/transistor, television, bicycle, scooter/ motorcycle/moped, none of the assets, computer/laptop); main source of lighting (electricity, kerosene, solar energy, other oil, any other, no lighting); Type of latrine facility within the premises (piped sewer system, septic tank, other system of flush latrine, with slab/ventilated improved pit, without slab/open pit, night soil disposed into open drain, night soil removed by human, night soil serviced by animal, public latrine, open latrine); Number of households having bathing facility within the premises (bathroom, enclosure without roof, no bathroom); Waste water outlet connected to (closed drainage, open drainage, no drainage); Cooking facility(kitchen cooked with fire-wood, crop residue, cow dung cake, coal/lignite/charcoal, kerosene, lpg/png, no kitchen but cooked with fire-wood, crop residue, cow dung cake, coal/lignite/charcoal, kerosene, lpg/png, no cooking). The domain is the districts of West Bengal as per Census 2011.

III. MATERIAL AND METHODS

11 main variables has been considered for analytical purpose in this study. These 11 main variables are consisting of several sub variables. Very few sub variables are excluded due to lack of importance, insignificant small data values and lesser impact in 2020 corresponding to 2010.

Method-1: Dataset obtained by excluding a very few sub variables in the above process and finally it consists of 79 variables. To identify the raw score whether it is equal to the average value or higher or lower than the mean average, the z_{ijk} values are calculated for each of the variables.

We define

$$x_{ijk} = j\text{th subvariable of } i\text{th main variable for the } k\text{th district.}$$

$$i = 1(1)11, \quad j = 1(1)79; \quad k = 1(1)19$$

$$z_{ijk} = (\max(x_{ijk}) - x_{ijk}) / (\max(x_{ijk}) - \min(x_{ijk}))$$

Principal Component Analysis has been done on these z_{ijk} 's. Almost 70% variations has been explained by the first factor. To compare the Eigen values from this result, Parallel Analysis has been done and as per this, the number of factors to be retained is 2. But since the first factor has explained 70% of variation, only the first factor has been considered for further analysis. The weights(w_{ijk} 's) have been taken as rotated varimax factor scores for the variables.

Index for a district is the weighted sum of weight's and z_{ijk} 's for the variables. The component values obtained from Rotated component matrix corresponding to first factor for each of the sub variables is considered as weightage value. The Index then defined as sum of the product of weight and z_{ijk} . Based on this Index, ranking has been made.

$$Index(k) = \sum \sum z_{ijk} * w_{ijk} \text{ over } i \& j$$

Since the index is calculated through z_{ijk} which is $(\max\text{-value})/(\max\text{-min})$, the smaller the index, the higher will be the ranking.

Method-2: In the second method, the sub variables are ranked according to their social importance, social infrastructure, economic value, and to some extent as a measure of social development. The highest value is assigned to the sub variable **within a main variable** which is having maximum social importance or best infrastructure quality, or economically high value, or having major contribution towards the development of the society and **finally it is** considered as weight. The sub-variables under each main variable are weighted according to their importance in social structure.

$$x_{ijk} = j\text{th subvariable of } i\text{th main variable for the } k\text{th district.}$$

$$i = 1(1)11, \quad j = 1(1)79; \quad k = 1(1)19$$

The raw score of each of this sub variables are transformed to Z variable as define below.

$$z_{ijk} = (\max(x_{ijk}) - x_{ijk}) / (\max(x_{ijk}) - \min(x_{ijk}))$$

$$w_{ijk} = \text{weightage of } z_{ijk}$$

The weighted Average of these z_{ijk} 's are then calculated using the weights already assigned as explained above.

$$\bar{z}_{ik} = \sum_j z_{ijk} * w_{ijk}$$

The value thus obtained represents a particular main variable corresponding to a particular district. Using the same technique, value of that particular variable is obtained for all Districts. Similar calculation has been done for all the 11 variables and obtained value of each variables for all the 19 Districts.

$$Y_1^{(1 \times 19)} = \bar{z}_{1k}$$

$$Y_2^{(1 \times 19)} = \bar{z}_{2k}$$

⋮
⋮
⋮

$$Y_{11}^{(1 \times 19)} = \bar{z}_{11k}$$

A matrix of 19x11 is thus obtained which is applied for Principal Component Analysis(PCA).

The component values obtained from the Component matrix corresponding to first factor for each of the variables is considered as weight and used to find the Index defined as sum of the product of this weight and the variables used for Principal Component Analysis from the 19x11 matrix.

$$Index = \sum Y_{ik} * c_i$$

Where c_i is the i^{th} component of the Rotated component matrix corresponding to first factor.

Table – 1 showing the ranks of the districts based on 1st method: Urban

Code	Rank	District	Index
327	10	Darjiling	43.82
328	9	Jalpaiguri	40.87
329	16	Koch Bihar	47.33
330	15	Uttar Dinajpur	46.93
331	19	DakshinDinajpur	48.09
332	13	Maldah	46.29
333	8	Murshidabad	40.41
334	14	Birbhum	46.51
335	4	Barddhaman	26.81
336	7	Nadia	37.07
337	1	North Twenty Four Parganas	5.53
338	5	Hugli	33.12
339	18	Bankura	47.7
340	17	Puruliya	47.42
341	3	Haora	25.95
342	2	Kolkata	12.57
343	6	South Twenty Four Parganas	33.92
344	11	PaschimMedinipur	44.52
345	12	PurbaMedinipur	45.02

Table – 2 showing the ranks of the districts based on 2nd method : Urban

Code	Rank	District	Index
327	10	Darjiling	9.72
328	9	Jalpaiguri	8.93
329	15	Koch Bihar	10.5
330	16	Uttar Dinajpur	10.5
331	19	DakshinDinajpur	10.7
332	13	Maldah	10.3

333	8	Murshidabad	8.8
334	14	Birbhum	10.3
335	3	Bardhaman	5.89
336	7	Nadia	8.07
337	1	North Twenty Four Parganas	1.52
338	6	Hugli	7.51
339	18	Bankura	10.7
340	17	Puruliya	10.6
341	4	Haora	5.95
342	2	Kolkata	3.63
343	5	South Twenty Four Parganas	7.44
344	11	PaschimMedinipur	9.91
345	12	PurbaMedinipur	9.98

To validate it with Statistical Tools, Wilcoxon Singed rank Test and Sign Test has been performed. There are 3 districts where ranks in 2nd method is higher than ranks in 1st method, 3 districts where the ranks in 1st method is higher than ranks in 2nd method and remaining 13 are having same rank in both the methods.

Table – 3 showing the ranks of the districts based on 1st method: Total

Code	Rank	District	Index
327	13	Darjiling	32.76
328	11	Jalpaiguri	30.08
329	16	Koch Bihar	33.72
330	18	Uttar Dinajpur	34.39
331	19	DakshinDinajpur	36.03
332	12	Maldah	32.31
333	7	Murshidabad	24.98
334	15	Birbhum	32.98
335	3	Bardhaman	16.86
336	8	Nadia	25.09
337	1	North Twenty Four Parganas	3.57
338	5	Hugli	20.66
339	14	Bankura	32.86
340	17	Puruliya	34.05
341	6	Haora	21.85
342	2	Kolkata	14.72
343	4	South Twenty Four Parganas	20.49
344	10	PaschimMedinipur	28.79
345	9	PurbaMedinipur	27.71

Table – 4 showing the ranks of the districts based on 2nd method: Total

Code	Rank	District	Index
327	17	Darjiling	9.33
328	11	Jalpaiguri	8.08
329	14	Koch Bihar	8.92
330	16	Uttar Dinajpur	9.28

331	19	DakshinDinajpur	9.97
332	12	Maldah	8.74
333	6	Murshidabad	6.47
334	13	Birbhum	8.83
335	2	Barddhaman	4.64
336	8	Nadia	6.87
337	1	North Twenty Four Parganas	1.97
338	5	Hugli	6.27
339	15	Bankura	8.95
340	18	Puruliya	9.46
341	7	Haora	6.84
342	4	Kolkata	5.68
343	3	South Twenty Four Parganas	5.42
344	9	PaschimMedinipur	7.35
345	10	PurbaMedinipur	7.46

To test the Rank statistically, here also, Wilcoxon Singed rank Test and Sign Test has been performed. There are 6 districts where ranks in 2nd method is lower than ranks in 1st method, 7districts where the ranks in 1stmethod is lower than ranks in 2ndmethod and remaining 6 are having same rank in both the methods.

REMARKS

Table – 5 showing the ranks for both the methods (URBAN)

District	Rank 1	Rank 2
Darjiling	10	10
Jalpaiguri	9	9
Koch Bihar	16	15
Uttar Dinajpur	15	16
DakshinDinajpur	19	19
Maldah	13	13
Murshidabad	8	8
Birbhum	14	14
Barddhaman	4	3
Nadia	7	7
North Twenty Four Parganas	1	1
Hugli	5	6
Bankura	18	18
Puruliya	17	17
Haora	3	4
Kolkata	2	2
South Twenty Four Parganas	6	5
PaschimMedinipur	11	11
PurbaMedinipur	12	12

The two ranks are so much similar, there are hardly anything which can be identified as a significant difference between the two ranks. Where there is a difference between ranks that is maximum of 1 point. So, these two approach perfectly matched the ranks and hence, the ranking thus obtained is very much significant.

Table – 6 showing the ranks for both the methods (TOTAL)

District	Rank 1	Rank 2
Darjiling	13	17
Jalpaiguri	11	11
Koch Bihar	16	14
Uttar Dinajpur	18	16
DakshinDinajpur	19	19
Maldah	12	12
Murshidabad	7	6
Birbhum	15	13
Barddhaman	3	2
Nadia	8	8
North Twenty Four Parganas	1	1
Hugli	5	5
Bankura	14	15
Puruliya	17	18
Haora	6	7
Kolkata	2	4
South Twenty Four Parganas	4	3
PaschimMedinipur	10	9
PurbaMedinipur	9	10

Ranking for both methods are very much similar, however, due to diversity in Rural and Urban frame, there are more untied cases here. Maximum deviation on ranking is seen in the District Darjeeling.

ACKNOWLEDGEMENT

Authors want to acknowledge

1. Office of the Registrar General, India for the Data, continuous encouragement to explore the Census data and for providing the scope of using STATA software with proper guidance.
2. Director of Census Operations, West Bengal.

Disclaimer: The findings of this paper is the view of the Authors. It has nothing to do to any Organization.

REFERENCE

- [1]. Ayaga, B and Tukufu, A (2004): Socioeconomic Status and Child Mortality: An Illustration, Using Housing and Household Characteristics from African Census Data, African Population Studies, Vol. 19, pp. 9-29
- [2]. Bhadra P P and Ghara T K (2020): Ranking Of Districts Of West Bengal By An Index Developed Through Household Data For Rural Areas of Census 2011 (communicated)
- [3]. Census 2011: Census of India 2011 Provisional Population Totals Urban Agglomerations and Cities : www.censusindia.gov.in/2011-prov-results/paper2/data_files/India2/1. Data Highlight.pdf
- [4]. Das, B and Mistri, A (2013): Household Quality of Living in Indian States: Analysis of 2011, Census: <https://doi.org/10.1177/0975425313477759>
- [5]. Directorate of Census Operators, West Bengal.
- [6]. Ghara, T K(2016): Classification of the States of India Based on Higher Education Development Indicators, IOSR Journal of Research & Method in Education, Volume 6, Issue 6 Ver. I, PP 65-70
- [7]. Office of the Registrar General, India. 2A ManSingh Road, New Delhi.
- [8]. Orr, D B (1964): The Importance of Weighting: An Illustrative Example, Educational and Psychological Measurement, VOL-XXIV, No.1. <https://journals.sagepub.com/doi/10.1177/001316446402400108>
- [9]. Panda, M(2014): Using Census Data to examine Rural Development – A Study of EAG States (a working paper)
- [10]. Sidak,Z, Sen, P K and Hajek, J (2004) : Theory of Rank Tests (Probability and Mathematical Statistics) 2nd Edition.
- [11]. Yongheng, Y, Angang, H and Ning, Z : An Alternative to Human Development Index with Principal Component Analysis (Center for National Condition Study, School of Public Policy &Management, Tsinghua University)
- [12]. http://en.cnki.com.cn/Article_en/CJFDTotal-JJYJ200507000.htm
- [13]. Data Source: <https://censusindia.gov.in/DigitalLibrary/TablesSeries2001.aspx>