ISSN 2348-3156 (Print)
International Journal of Social Science and Humanities Research ISSN 2348-3164 (online)
Vol. 9, Issue 3, pp: (332-338), Month: July - September 2021, Available at: www.researchpublish.com

EVALUATION OF DISTRICTS OF NORTH EASTERN STATES OF INDIA BY AN INDEX DEVELOPED THROUGH PROBABILISTIC APPROACH BASED ON HOUSEHOLD DATA FOR RURAL AREAS OF CENSUS 2011

Partho Pratim Bhadra[1], Tushar Kanti Ghara[2]

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

[2] Joint Director of Public Instruction, Department of Higher Education, Govt. of West Bengal, 8th Floor, Bikash Bhavan, Salt Lake City, Kolkata 700091

Abstract: **Population Census in India is conducted once in every ten years under the census Act 1948 and Census rules 1990. Census of India is conducted in two phases. First phase is House listing and Housing Census and second phase in Population Enumeration. Here an attempt has been made to evaluate the districts of all eight states of North East India on the basis of Rural Data of House listing and Housing Census of Census of India 2011. The three broad categories of House listing & Housing Census data are Housing condition, Amenities and Assets, based on which the analysis has been done. Statistical techniques like Probability, Rank Test and other Descriptive Statistical Theories have been applied.**

Keywords: **Probability, Rank Test, Arithmetic mean, Geometric mean.**

1. INTRODUCTION

The states located on the north east part of India are regarded as the North-East India States. These north-east Indian states include Sikkim and the Seven Sister States of India, which are, Assam, Arunachal Pradesh, Manipur, Meghalaya , Mizoram , Tripura and Nagaland .

Location of the north eastern India makes it share its border with numerous countries like Bhutan, China and Nepal in the north, Bangladesh in the west and Burma in the east and south.

India is a welfare State. Since independence, various welfare schemes have been launched for the welfare of the common man. This information is provided by the Census. 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. The Houselisting and Housing Census has immense utility as it provides comprehensive data on the conditions of human settlements, housing deficit and consequently the housing requirement to be taken care of in the formulation of housing policies. Bhadra and Ghara (2020).

This also provide a wide range of data on amenities and assets available to the households, information much needed by various departments of the Union and State Governments and other non-Governmental agencies for development and planning at the local level as well as the State level. This also provide the base for Population Enumeration. For Census

purposes, total geographical area is broadly classified into Rural & Urban. The basic Unit of rural area is revenue Village (CENSUS OF INDIA, 2011). The rural data set is considered here for the evaluation of the Districts across states. Household quality of living (HQL) refers to three broad aspects including housing condition, amenities and assets. The study is an attempt to investigate regional variation across the districts of North East Indian States based on Rural data considering the Housing infrastructure, Availability of Household Amenities and Assets. Using the Classical probability approach and then adopting the Weighted Arithmetic Mean and Weighted Geometric Mean the analytical process have been done by using an Index, which is further tested by Correlation and Rank Test. The data analysis has been done by using SPSS and Excel.

2. DATA

The main categories of variables are Housing infrastructure, Availability of Household Amenities and Assets. We consider here 11sub categories under which 85 variables are considered. The study is based on 85 such variables.

1. **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);

2. **Condition of census houses** (good, livable, dilapidated);

3. **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);

4. **Material of wall** (grass/ thatch/ bamboo/ wood/ mud/etc, plastic/ polythene, Mud/ Unburnt brick, Wood, Stone not packed with mortar, Stone packed with mortar, GI/metal/asbestos sheets, Burnt brick, concrete, any other material);

5. **Material of floor** (mud, wood/ bamboo, burnt brick, stone, cement, mosaic/floor tiles, any other material);

6. **Availability of assets** (radio/ transistor, television, computer/ laptop , bicycle, scooter/ motorcycle/moped, none of the assets,);

7. **Main source of lighting** (electricity, kerosene, solar energy, other oil, any other, no lighting);

8. **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 of Pit Latrine, night soil disposed into open drain, night soil removed by human, night soil serviced by animal, *No latrine within premises* (public latrine, open latrine);

9. **Number of households having bathing facility** within the premises (bathroom, enclosure without roof) and no bathroom;

10. **Waste water outlet connected to**(closed drainage, open drainage, no drainage);

11. **Cooking facility** (***Has kitchen*** - fire-wood, crop residue, cow dung cake, coal/lignite/charcoal, kerosene, lpg /png, Electricity, Biogas ,Any other), (***does not have kitchen*** -fire-wood, crop residue, cow dung cake, coal/lignite/charcoal, kerosene, lpg /png, Electricity, Biogas, Any other) and ***no cooking***.

Observation on the data

Two main approaches those are used for the analysis are Arithmetic Mean and Geometric Mean and assigning weight in each procedure. It has been observed while applying the Index for weighted G.M., there are several variables for which the data returns the value zero (0).

These variables are :Uses of

(i)cow dung cake, (ii) coal/lignite/charcoal, (iii) kerosene, (iv) lpg /png,, (v) Electricity, (vi) Biogas under sub category “Does not have kitchen” and (i) cow dung cake, (ii) coal/lignite/charcoal, (iii) kerosene, (iv) Electricity, (v) Biogas, (vi) Any other under sub category “Has Kitchen”, which comprise 12 variables.

Other three variables for which there are zero (0) values in several districts are (i) Night soil disposed into open drain,(ii)Night soil removed by human and (iii) Night soil serviced by animal.

ISSN 2348-3156 (Print)
International Journal of Social Science and Humanities Research ISSN 2348-3164 (online)
Vol. 9, Issue 3, pp: (332-338), Month: July - September 2021, Available at: www.researchpublish.com

3. MATERIAL AND METHOD

85 variables have been considered for analytical purpose in this study. Very few variables are excluded due to lack of importance, insignificant small data values and lesser impact in 2020 in respect to 2010.

$$X_{ijk} = Value\ of\ j^{th}\ subvariable\ of\ i^{th}\ main\ variable\ for\ the\ k^{th}\ district.$$

$$i = 1(1)11;\ j = 1(1)n(i)\ ;\ \ k = 1(1)18$$

We define for fixed i and k,

$$\boldsymbol{P_{ijk} = X_{ijk} / \sum_{j=1}^{n(i)} X_{ijk}}$$

$$i = 1(1)11,\ \ ;\ \ k = 1(1)18, j = 1(1)n(i) \text{number of sub variables in ith subcategory}$$

1st approach:

Considering
$w_{ijk} = weightage\ of\ P_{ijk}$ = 1/N, where N= sum of n(i) = 85

Using Weighted Arithmetic Mean procedure, we define

$$Index(I1) = \frac{\sum P_{ijk}^{W_{ijk}}}{\sum P_{ijk}^{W_{ijk}} + \sum (1-P_{ijk})^{W_{ijk}}} \qquad \text{.....(1)}$$

2nd approach :

$w_{ijk} = weightage\ of\ P_{ijk}$ = 1/N, where N= sum of n(i) = 85

Using Weighted Geometric Mean procedure, we define

$$I2 = \frac{\prod P_{ijk}^{W_{ijk}(p)}}{\prod P_{ijk}^{W_{ijk}(p)} + \prod (1-P_{ijk})^{W_{ijk}(p)}} \qquad \text{.....(2)}$$

Since overall there are fifteen variables for which we have 0 value and the above said Index will lead to a value which is also 0.

Therefore instead of using the above (2), a more generalize formula is being used.

To deal with a data set with zeros a process that is adopted here is to add one to each value in the data set, calculating the geometric mean of this shifted data set and then subtracting one again:

Based on the Modified Weighted G.M., we define the Index as below.

$$Index\ (I2) = \frac{exp(\frac{1}{\sum Wijk}\sum_i (\text{Wijk}) \log(Pijk + 1)) - 1}{(exp(\frac{1}{\sum Wijk}\sum_i (\text{Wijk}) \log(Pijk + 1)) - 1 + exp(\frac{1}{\sum Wijk}\sum_i (\text{Wijk}) \log(1 - Pijk + 1)) - 1}$$

However since all the weights are equal, the weighted geometric mean is the same as the *geometric mean. So, finally we will use*

$$Index\ (I2_{Corrected}) = \frac{exp(\frac{1}{n}\sum_i \log(Pijk + 1)) - 1}{(exp\left(\frac{1}{n}\sum_i \log(Pijk + 1)\right) - 1) + (exp(\frac{1}{n}\sum_i \log(1 - Pijk + 1)) - 1)}$$

ISSN 2348-3156 (Print)
International Journal of Social Science and Humanities Research ISSN 2348-3164 (online)
Vol. 9, Issue 3, pp: (332-338), Month: July - September 2021, Available at: www.researchpublish.com

Using Index(I1) and Index (I2 corrected) the following table is prepared.

Table 1: showing the ranks of the districts based on both methods.

State	District	Index (I1)		Index (I2 Corrected)	
		Index value	Rank 1	Index value	Rank 2
SIKKIM	North District	0.001736	83	0.116432	63
	West District	0.001735	84	0.115281	72
	South District	0.001743	78	0.116173	65
	East District	0.001762	62	0.117468	43
ARUNACHAL PRADESH	Tawang	0.001766	54	0.117073	53
	West Kameng	0.001767	51	0.118112	37
	East Kameng	0.001762	61	0.114860	78
	Papum Pare	0.001841	16	0.123667	8
	Upper Subansiri	0.001766	58	0.115257	73
	West Siang	0.001801	36	0.120273	23
	East Siang	0.001850	12	0.122463	15
	Upper Siang	0.001766	57	0.116873	54
	Changlang	0.001820	25	0.120150	24
	Tirap	0.001753	66	0.113958	82
	Lower Subansiri	0.001802	35	0.120623	20
	Kurung Kumey	0.001748	72	0.112577	85
	Dibang Valley	0.001770	49	0.119118	28
	Lower Dibang Valley	0.001832	20	0.122874	12
	Lohit	0.001828	23	0.120906	19
	Anjaw	0.001754	65	0.116052	66
NAGALAND	Mon	0.001745	74	0.113870	83
	Mokokchung	0.001786	41	0.116835	56
	Zunheboto	0.001752	69	0.114916	77
	Wokha	0.001781	44	0.118522	33
	Dimapur	0.001827	24	0.121558	17
	Phek	0.001742	79	0.114354	80
	Tuensang	0.001737	81	0.115158	76
	Longleng	0.001740	80	0.114027	81
	Kiphire	0.001732	85	0.112244	86
	Kohima	0.001775	46	0.116853	55
	Peren	0.001745	75	0.115819	67
MANIPUR	Senapati	0.001769	50	0.117524	41
	Tamenglong	0.001762	63	0.116738	58
	Churachandpur	0.001818	27	0.123383	10
	Bishnupur	0.001914	3	0.124951	3
	Thoubal	0.001910	4	0.124493	4
	Imphal West	0.001983	1	0.129605	1
	Imphal East	0.001950	2	0.127832	2
	Ukhrul	0.001763	59	0.116753	57
	Chandel	0.001828	22	0.121950	16
MIZORAM	Mamit	0.001766	55	0.117886	39
	Kolasib	0.001749	70	0.117516	42

ISSN 2348-3156 (Print)
International Journal of Social Science and Humanities Research ISSN 2348-3164 (online)
Vol. 9, Issue 3, pp: (332-338), Month: July - September 2021, Available at: www.researchpublish.com

	Aizawl	0.001783	42	0.117389	47
	Champhai	0.001762	60	0.116466	61
	Serchhip	0.001766	56	0.115195	75
	Lunglei	0.001767	52	0.117303	48
	Lawngtlai	0.001755	64	0.115257	74
	Saiha	0.001745	76	0.115715	68
TRIPURA	West Tripura	0.001817	28	0.118326	36
	South Tripura	0.001799	37	0.117166	50
	Dhalai	0.001777	45	0.117110	52
	North Tripura	0.001770	47	0.117175	49
MEGHALAYA	West Garo Hills	0.001795	39	0.117949	38
	East Garo Hills	0.001805	33	0.117428	45
	South Garo Hills	0.001783	43	0.117423	46
	West Khasi Hills	0.001749	71	0.116185	64
	Ribhoi	0.001767	53	0.118525	32
	East Khasi Hills	0.001746	73	0.116607	59
	Jaintia Hills	0.001729	86	0.116539	60
ASSAM	Kokrajhar	0.001790	40	0.115583	69
	Dhubri	0.001753	67	0.113429	84
	Goalpara	0.001812	30	0.118427	35
	Barpeta	0.001797	38	0.117450	44
	Morigaon	0.001813	29	0.118584	31
	Nagaon	0.001803	34	0.118774	29
	Sonitpur	0.001831	21	0.120360	22
	Lakhimpur	0.001838	17	0.120366	21
	Dhemaji	0.001835	18	0.119747	26
	Tinsukia	0.001846	14	0.122572	14
	Dibrugarh	0.001871	8	0.123968	7
	Sivasagar	0.001872	5	0.124348	6
	Jorhat	0.001869	9	0.123608	9
	Golaghat	0.001854	11	0.121463	18
	Karbi Anglong	0.001834	19	0.119223	27
	Dima Hasao	0.001752	68	0.114751	79
	Cachar	0.001770	48	0.117545	40
	Karimganj	0.001737	82	0.115282	71
	Hailakandi	0.001745	77	0.115330	70
	Bongaigaon	0.001806	32	0.117139	51
	Chirang	0.001806	31	0.116439	62
	Kamrup	0.001862	10	0.122622	13
	Kamrup Metropolitan	0.001872	6	0.124356	5
	Nalbari	0.001871	7	0.123063	11
	Baksa	0.001844	15	0.118453	34
	Darrang	0.001820	26	0.118587	30
	Udalguri	0.001850	13	0.119993	25

Here Rank is basically the serial number in ascending order of Index value.

To validate it with Statistical Tools, Wilcoxon Singed rank Test has been performed.

ISSN 2348-3156 (Print)
International Journal of Social Science and Humanities Research ISSN 2348-3164 (online)
Vol. 9, Issue 3, pp: (332-338), Month: July - September 2021, Available at: www.researchpublish.com

Ranks				
		N	Mean Rank	Sum of Ranks
VAR00003 - VAR00002	Negative Ranks	38[a]	42.46	1613.50
	Positive Ranks	42[b]	38.73	1626.50
	Ties	6[c]		
	Total	86		
a. Rank 2 < Rank1				
b. Rank 2 >Rank 1				
c. Rank 2 = Rank 1				

There are 38 districts where ranks in 2nd method is lower than ranks in 1st method, 42 districts where the ranks in 2nd method is higher than ranks in 1st method.

6 cases where we have same ranks.

Test Statistics[a]	
Z-value	-0.031[b]
Asymp. Sig. (2-tailed)	0.975
a. Wilcoxon Signed Ranks Test	
b. Based on negative ranks.	

As per result of The Wilcoxon Singed rank Test Z value is - 0.032 with p=0.975. The two tailed test shows the test of Rank is statistically very much significant.

4. REMARKS

The ranks are grouped into three clusters. Considering ranks obtained from Index (I1), three groups formed. Class boundaries are (i) 1-28 (ii)29-57 and (iii) 58-86 based on Index (I1). Then a checking is done to see whether a district having rank between 1-28 by Index (I1) also have rank between 1-28 by Index(I2 Corrected) and so on for other groups.

Group 1: out of 28 districts, 25 will fall in the group for both Index. 3 districts having rank between 1-28 by Index (I1) is not in this group according to Index (I2 Corrected)

The three districts areWest Tripura of Tripura, Baksa and Darrang of Assam. Mathematically, they might not fall in the group of 1-28 but they are extremely close to the group according to the rank Index (I2 Corrected)

It may be observed that Ranks obtained by 2nd Index is at the borderline and not that much distant from 1st Index.

Group 2: (Rank between 29 to 57) out of 29 districts 24 districts having rank between 29-57 obtained from both Index.5 districts differ. These districts are West Siang, Lower Subansiri and Dibang Valley of Arunachal Pradesh, Serchhip district of Mizoram and Kokrajhar of Assam. Among thee, District - Dibang Valley is on extreme borderline (rank 28 for 2nd Index) and , Serchhip district of Mizoram is slightly distant.Others are also close to group.

Group 3: (Rank between 58-86) Here , out of 29 districts, only 2 districts are therenamely, East District of Sikkim and District – Kolasib of Mizoram.

5. ACKNOWLEDGEMENT

Office of the Registrar General, India and Director of Census Operations, West Bengal is being acknowledged for using data for research works only.

DISCLAIMER

The findings of this paper are the views of the Authors. It has nothing to do to any Organization.

REFERENCES

[1] 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 (accepted IJRSR)

[2] Bhadra, P P and Ghara, T K (2020): Ranking of Districts of West Bengal by an Index Developed Through Census 2011 Household Data (Urban & Total), International Journal of Humanities and Social Science Invention (IJHSSI), Vol-9 No-7 Pp31-36

[3] Census 2011: Census of India 2011: www. censusindia.gov.in

[4] Das, B and Mistri, A (2013): Household Quality of Living in Indian States: Analysis of 2011, Census.

[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] Jian L, Saha, B and Deshpande, A(2009) : A Unified Approach to Ranking in Probabilistic Databases

[8] Leclerc, G, Nelson, A and Knapp, E B(2000): The use of unit-level census data for research on poverty: a multiscale approach; International Center for Tropical Agriculture (CIAT), Colombia.

[9] Roberto de la Cruza, Jan-Ulrich Krefta: Geometric mean extension for data sets with zeros

[10] Central limit theorem, Encyclopedia of Mathematics, Springer Netherlands.