International Journal of Advanced Community Medicine

E-ISSN: 2616-3594
P-ISSN: 2616-3586
IJACM 2019; 2(3): 33-36
Received: 21-07-2019
Accepted: 25-08-2019

Deepak Jamadar
Assistant Professor cum Statistician, Department of Community Medicine, SVS Medical College, Yenugonda, Mahabubnagar, Telangana, India

Dr. KP Joshi
Professor, Department of Community Medicine, SVS Medical College, Yenugonda, Mahabubnagar, Telangana, India

Correspondence
Dr. KP Joshi
Professor, Department of Community Medicine, SVS Medical College, Yenugonda, Mahabubnagar, Telangana, India

Statistical analysis of urbanization and an application towards a Pareto model

Deepak Jamadar and Dr. KP Joshi

DOI: https://doi.org/10.33545/comed.2019.v2.i3a.81

Abstract
Background: India is only 23.7 percent urban at the present time and this cannot be considered a high level of urbanization. However, when we look at the number of people living in the cities (around 156 million out of the total population of 684 million in 1981) the significance of urban population assumes importance.
Aims and objective: To fit Pareto model to the urban population data of Tamil Nadu state and to compare the urbanization in Tamil Nadu over the 3 decades (1981, 1991, 2001).
Materials and method: we were collected three years Indian census population data of Tamil Nadu state (1981, 1991, and 2001) As city size models are in skew nature the Pareto model has been fitted to the Tamil Nadu population data for the above said three years and the fitness of the model were tested using chi-square statistics.
Result: The estimates of the parameters a, k for the years 1981, 1991 and 2001 are 1.784, 6039.4224, 1.3901, 5807.60 and 1.79572, 6688.082, the χ^2 –statistics for the years 1981, 1991 and 2001 are 171.64, 104.737 and 304.819, i.e., for all the three years the $\chi^2_{0.05} < \chi^2_{\text{calculated value}}$ hence the Pareto model is not good fit for the all the population data.
Conclusion: The city size distribution is skewed distributions. The less variation between the observed and expected number of cities in all the 3 years. Thus the Pareto model may be proposed and we found that the Pareto model is not a good fit for all the population data.

Keywords: Pareto model, urbanization, population, model

Introduction
India's level of urbanization increased from 17.6 per cent in 1951 to only 23.7 per cent in 1981 and 27.8 per cent in 2001 [1]. However, when we look at the number of people living in the cities (around 156 million out of the total population of 684 million in 1981) the significance of urban population assumes importance. In the west, urbanization accompanied migratory movements from rural to urban areas and was generally considered desirable and provided employment opportunities and other facilities. It was considered a symbol of man's mastery over nature and thus a symptom of progress. In countries like India on the other hand, it represents a transfer of rural poverty to the urban areas and we find that in some of the metropolitan cities the number of slum dwellers' quite significant. In fact, some scholars have indicated that in countries like India what is happing in urbanization without urban development leading to what is commonly called "Over urbanization" It means that the Indian cities are growing more rapidly than the capacity of the economy as a whole to achieve substantial net gains of self-generating type which would allow sufficient margin for social investment. The important point that should be emphasized here is that by comparison with other developing region of the world, the growth rate of the urban population in India is noticeably slow, but the scale of urban population is increasing sharply in each decade. In other words, despite a sluggish growth of urbanization in India, the scare of population pressure on the existing cities and towns in accelerating rather sharply. Tamil Nadu State is situated at the South Eastern extremity of the Indian Peninsula bounded on the north by Karnataka and Andhra Pradesh on the east by Bay of Bengal, on the South by the Indian Ocean and on the West by Kerala State.
Tamil Nadu ranks first in urbanization among the fifteen major States in the country. Urbanization has been on the increase since 1901. According to the 2001 Census, Tamil Nadu has emerged as the State with the highest level of urbanization (44%) in the country

among the larger States. While the percentage of urban population in the country increased from 10.85% to 27.78% during 1901-2001, Tamil Nadu registered a much higher percentage increase i.e., from 14.15% to 44.04% during the above period. With the help of above facts and figures we made an effort to find, whether pareto model is good fit or not in changing pattern of urbanization in Tamil Nadu state.

Materials and Methods

To study the urbanization of Tamil Nadu state we were collected three years Indian census population data of Tamil Nadu state (1981, 1991, and 2001) the City size is defined as the number of population in the city, City size distribution in the distribution of cities with respect to the size. Changes the city sizes are due to random factors. Hence city sizes are also called as random variables. The distribution of the frequencies in the different classes is called a frequency table and the manner in which the class frequencies are distributed over the class intervals is called the grouped frequency distribution of the variable. It is used to form the city size distribution. Having fixed the number of classes, divide the range (the difference between the greatest and smallest observation) by it and the nearest integer to this values gives the magnitude of the class interval. Broad class intervals [(i.e), less number of classes] will yield only rough estimates while for high degree of accuracy small class intervals [(i.e.,) large number of classes] are desirable. The class limits should be chosen in such a way that the mid-value of the class interval and actual average of the observation in that class interval are as near to each other as possible. If this is not the case then the classification gives a disported picture of the characteristic of the data. If possible, class limits should be located at the points which are multiple of 0,2,5,10,…..etc., so that the midpoints of the classes are the common figures *viz*., 0,2,5,10,……etc., the figures capable of easy and simple analysis.

Pareto model

Statistical model are the models connecting the random variables by means of statistical statements. Model such as lognormal, Pareto, rank size etc., are use to describe the urbanization of any area. As city models are in skew nature, skew models may be proposed to study the skewed city size distribution. Skewed distributions are log normal, Pareto, rank size etc., to study the city size pattern Pareto model may be proposed.

The Pareto distribution of the third kind has the cumulative distribution function

$$F_x(X) = 1 - k_2 e^{-bx} / (x + 1)^a$$

We will consider only distribution.

The Pareto density function is,

$P_{x(}x) = ak /x^a + 1, (a>0, x \geq k>0)$

Results

Table 1. Shows City size and Number of cities

City size	Number of cities
City size distribution - 1981	
0-5000	18
5000-10000	65
10000-20000	102
20000-50000	81
50000-100000	35
100000-500000	19
City size distribution - 1991	
0-5000	3
5000-10000	49
10000-20000	113
20000-50000	91
50000-100000	41
100000-500000	23
City size distribution - 2001	
0-5000	4
5000-10000	36
10000-20000	96
20000-50000	107
50000-100000	53
100000-500000	22

Table 2: Estimates of the Parameters computation for the year 1981

X	Mid-point of x	F	P	Observed cumulative F	1-F	Y=log(1-F)	$X=\log_e x$
0-5000	2500	18	0.05625	0.05625	0.94375	-0.0579	7.8240
5000-10000	7500	65	0.20313	0.25938	0.74062	-0.3003	8.9227
10000-20000	15000	102	0.31875	0.57813	0.42187	-0.8631	9.6158
20000-50000	35000	81	0.25313	0.83126	0.16874	-1.7794	10.4631
50000-10000	75000	35	0.10934	0.9406	0.0594	-2.8235	11.2252
10000-500000	300000	19	0.05938	0.9999	0.0001	-9.2103	12.6115

Table 3: computation of expected number of cities

X	$\hat{Y}=A+BX$	$e^y = 1 - F$	$\hat{F}$	$\hat{P}$	$\hat{N}\hat{P} = E$
7.8240	1.5739	4.8254	-3.8254	-	-
8.9227	-0.3866	0.6794	0.3206	0.3206	102.592≈103
9.6158	-1.6233	0.1972	0.8028	0.4822	154.304≈154
10.2252	-3.1353	0.0435	0.9565	0.1537	49.184≈49
11.2252	-4.4951	0.0112	0.9888	0.0323	10.336≈10
12.615	-6.9689	0.0009	0.9991	0.0103	3.296≈3
					320

Table 4: observed and expected number of cities

City size	Observed Frequency (O)	Expected Frequency (E)	$(O-E)^2 / E$
0-5000	18	-	3.883
5000-10000	} 65	} 103	
10000-20000	102	154	17.558
20000-50000	81	49	20.898
50000-100000	35	10	129.308
100000-500000	} 19	} 3	
			171.647

Table 5: Estimates of the Parameters computation for the year 1991

X	Mid-point of x	F	P	Observed cumulative F	1-F	Y=log(1-F)	$X=\log_e x$
0-5000	2500	3	0.009	0.009	0.991	-0.0090	7.8240
5000-10000	7500	49	0.153	0.162	0.838	-0.1767	8.9227
10000-20000	15000	113	0.353	0.515	0.485	-0.7236	9.6158
20000-50000	35000	91	0.284	0.799	0.201	-1.6045	10.4631
50000-10000	75000	41	0.128	0.927	0.073	-2.6173	11.2252
10000-500000	300000	23	0.072	0.999	0.001	-6.9078	12.6115

Table 6: computation of expected number of cities

X	$\hat{Y}=A+BX$	$e^y = 1 - F$	$\hat{F}$	$\hat{P}$	$\hat{N}\hat{P} = E$
7.8240	1.1718	3.2277	-2.2277	-	-
8.9227	-0.3555	0.70008	0.2992	0.2992	96
9.6158	-1.3190	0.2674	0.7326	0.4334	139
10.2252	-2.4969	0.0823	0.9177	0.1851	60
11.2252	-3.5563	0.0285	0.9715	0.0538	17
12.615	-5.4833	0.0042	0.9958	0.0243	8
					320

Table 7: observed and expected number of cities

City size	Observed Frequency (O)	Expected Frequency (E)	$(O-E)^2 / E$
0-5000	3	-	20.167
5000-10000	}	}	
	49	96	
10000-20000	113	140	5.207
20000-50000	91	59	17.356
50000-100000	41	17	33.882
100000-500000	23	8	28.125
			104.737

Table 8: Estimates of the Parameters computation for the year 2001

X	Mid-point of x	F	P	Observed cumulative F	1-F	Y=log(1-F)	$X=\log_e x$
0-5000	2500	4	0.01256	0.01256	0.98744	-0.01264	7.8240
5000-10000	7500	36	0.11321	0.12577	0.87423	-0.13441	8.9227
10000-20000	15000	96	0.30189	0.42766	0.57234	-0.55802	9.6158
20000-50000	35000	107	0.33648	0.76414	0.23586	-1.44452	10.4631
50000-10000	75000	53	0.16667	0.93081	0.06919	-2.67089	11.2252
10000-500000	300000	22	0.6919	0.9999	0.0001	-9.21034	12.6115

Table 9: computation of expected number of cities

X	$\hat{Y}=A+BX$	$e^y = 1 - F$	$\hat{F}$	$\hat{P}$	$\hat{N}\hat{P} = E$
7.8240	1.76174	5.85406	-4.85406	-	-
8.9227	-0.20582	0.81398	0.18602	0.18602	59
9.6158	-1.45043	0.23447	0.76053	0.57951	184
10.2252	-2.97195	0.5120	0.9488	0.18327	59
11.2252	-4.34047	0.01303	0.98697	0.0817	12
12.615	-6.82987	0.001	0.9989	0.1193	4
					318

Table 10: observed and expected number of cities

City size	Observed Frequency (O)	Expected Frequency (E)	$(O-E)^2 / E$
0-5000	4	-	6.119
5000-10000	}	}	
	36	59	
10000-20000	96	184	42.087
20000-50000	107	59	39.051
50000-100000	53	12	217.563
100000-500000	}	}	
	22	4	
			304.819

Table 11: Estimation of the parameters

Year	Parameters	Estimate
1981	a	$\hat{a} = 1.784$
	k	$\hat{K} = 6039.4224$
1991	a	$\hat{a} = 1.3901$
	k	$\hat{K} = 5807.60$
2001	a	$\hat{a} = 1.79572$
	k	$\hat{K} = 6688.082$

Table 12: Chi-square-values

Year	Parameters	Estimate
1981	$\chi^2 = 171.644$	$\chi^2_{(0.05)} < \chi^2$ calculated value 7.815 <171.644 The Pareto Model is not a good fit
1991	$\chi^2 = 104.737$	$\chi^2_{(0.05)} < \chi^2$ calculated value 9.488 <104.373 The Pareto Model is not a good fit
2001	$\chi^2 = 304.819$	$\chi^2_{(0.05)} < \chi^2$ calculated value 7.815 <304.819 The Pareto Model is not a good fit

Pareto model has been fitted to the Tamil Nadu population data of 1981, 1991 and 2001. The estimates of the parameters were obtained. The fitness of the model were tested using χ^2 –statistics and the inference were obtained. The average city size for the year 1981, 1991, 2001 were obtained

4. Discussion

Since independence the Urbanisation in India has been steadily increasing the rate of urbanisation has been far slower than other countries at similar stages of development. From 1901 to 2001, it can be witnessed a gradual increase in the urbanisation levels and growth of urban population in Tamil Nadu, comparatively higher than the all India's level. Tamil Nadu ranks first in urbanization among the fifteen major States in the country. Urbanization has been on the increase since 1901. According to the 2001 Census, Tamil Nadu has emerged as the State with the highest level of urbanization (44%) in the country among the larger States. While the percentage of urban population in the country increased from 10.85% to 27.78% during 1901-2001, Tamil Nadu registered a much higher percentage increase i.e., from 14.15% to 44.04% during the above period. With above facts and figures we made an effort to fit a pareto model whether it is good fit to three years population data in the changing pattern of urbanization in Tamil Nadu state.

5. Conclusion

In this study we found that, the number of cities is less than the number of cities in 1991 and 2001, and the city size distribution is skewed distributions. The less variation between the observed and expected number of cities in all the 3 years. Thus the Pareto model may be proposed to study the skewed distributions through its goodness is not good. Urban populations have been increased due to the growth of industries employment opportunities etc., in urban areas and in the present study we also found that, the Pareto model is not a good fit for all the three years i.e., 1981, 1991 and 2001.

References

1. Rakesh Mohan, Shubhagato Dasgupta. Urban Development in India in the Twenty First Century: Policies for Accelerating Urban Growth. Stanford Centre for International Development on, 2004.
2. Bhaskar D. Misra: An Introduction to the Study of Population, South Asian Publishers, 1982.
3. Census of India series -20 Tamil Nadu Part II-A. 1981
4. Census of India series -23 Tamil Nadu Provisional Population Total. 1991.
5. Census of India series -34 Tamil Nadu Provisional Population Total, Director of Census Population Total. 2001.
6. Director of Census operation, Tamil Nadu, Chennai, Primary Census abstract Census. 2001
7. Gupta SC, Kapoor VK. Fundamentals of Mathematical Statistics, Sultan Chand and Sons, New Delhi, 2003.
8. Rohatgi VK. An Introduction to Probability Theory and Mathematical Statistics. John Wiley & Sons, New York London Sydney Toronto, 1976.