Volume : 4 | Issue : 4 | April 2015 • ISSN No 2277 - 8179

Research Paper

Empirical study on Rural Taluk Size Distribution of Population data in Tamil Nadu state

Statistics

KEYWORDS :

Dr. B. Renganathan — Reader and Head (Rtd), Department of Statistics, Salem Sowdeswari College, Salem-10, Tamil Nadu, India

Mrs. M.Kalaiarasi — Assistant Professor of Statistics, Sri Ganesh College of arts and science, Ammapet, Salem-14, Tamil Nadu, India.

ABSTRACT

Ruralisation is the process of growth with respect to the area and population. Rural areas are the geographical regions which are located out of the town or cities. The average rural populations in 2001 are observed as 176000 and in 2011 as 175000.It shows the decreasing nature of rural population in TAMILNADU State. Similarly the rural population in India 2001 is 102.9 crores and 121crores in 2011. It shows the growth of the rural population in India. Thus the growth of the rural population in India as well as in TAMILNADU motivates to carry out the statistical study on ruralisation in TAMILNADU state. Statistical models such as exponential, q-exponential, truncated exponential and lognormal models are applied to study the nature of the rural taluk size distribution of population using 2001 and 2011 population census data of TAMILNADU state. The suitability of these four models is to be tested empirically using statistical test. Lognormal model is the more appropriate model than the other models for the taluk rural size population data.

1. Introduction

The populations of the region are dichotomized with respect to the residential status such as rural and urban. Rural areas are the geographical locations which are located out from the towns or cities. It has very low density of population. It may be less than 5000 people and it is the lowest unit of population. Agriculture is the main source of livelihood along with fishing, cottege industries, poultry farms, etc. in the villages. The aim of rural development is mainly to improve the social and economical level of rural areas. Rural development aims to find the ways of improving rural people themselves. The revenue administrative units are described as District, Taluk, and Villages in TAMILNADU State. L.c.Malacarne and R.S.Mendes (2001) applied the q-exponential distribution for describing the distribution of cities for all ranges of population. Keith Briggs (2006) studied the distribution of train delays on the British railway network using q-exponential model. González-Val, Rafael and Ramos, Arturo and Sanz, Fernando and Vera-Cabello, Maria (2013) applied lognormal, q-exponential, double parato and lognormal models to study the city size distributions. None of the above said papers deal with the study of rural taluk size distribution in TAMILNADU state.

Rural population –India Ruralisation to in India (2001-2011) Population (in crores)

	2001	2011	Difference
India	102.9	121	18.1
Rural	74.3	83.3	9.0
Urban	28.6	37.7	9.1

Of the 121 crore Indians, 83.3 crore live in rural areas while 37.7 crore stay in Urban areas. The rural population as well as urban population are having increasing tendency over the span of ten years at India level.

Rural population-Tamilnadu Ruralisation in Tamilnadu (2001-2011) Population (in crores)

	2001	2011	Difference
Tamilnadu	6.24	7.21	0.97
Rural	3.49	3.72	0.23
Urban	2.75	3.49	0.74

According to 2011 census, of the 7.21 crore Tamilnadu population, 3.72 crore live in rural areas while 3.49 crore stay in Urban areas. The rural population as well as urban population are having increasing tendency over the span of ten years at Tamilnadu level.

It motivates to propose the empirical study on rural taluk size distribution of TAMILNADU population using statistical models such as exponential, q-exponential, truncated exponential and lognormal model.

2. Objectives

(i) General objective

To study the growth nature of rural taluk size population of TAMILNADU state using the statistical models such as exponential, q-exponential, truncated exponential and lognormal model.

(ii) Specific objective

To test the suitability of these four statistical models using 2001 and 2011 census population data of TAMILNADU state.

3. Data Source

Census population 2001 and 2011 data are applied as a secondary data for analyzing the nature of taluk size distribution of TAMILNADU state.

4. Methodology

Definition

Variable

Number of population living in rural taluk is defined as rural taluk size. It is referred as a random variable because changes in rural taluk size population are random in nature.

Rural area

An area outside of the limits of a city or town or a designated commercial, industrial, or residential center is called as rural area. Rural areas are characterized by farms, vegetation and open spaces. Farmlands, homes and buildings are spread out with larger distance. Rural is the natural environment and the opposite of the city life and pertaining to the less-populated areas.

Taluk size distribution

Taluk size is the no. of population living in the taluk. Taluk size distribution is the distribution of Taluk with respect to their population.

Models

A model establishes the main variables involved and connects them by means of mathematical statements. Thus, a model simplifies the reality and points out the basic features of the problem and denotes a set of hypothesis.

5. Methods

5.1 Method of Moments

$X_1, X_2, \ldots X_n$ be a random sample of size n drawn from a given population having the probability density function $f(x;\theta 1,\theta 2\ldots,\theta k)$ with k parameters $\theta_1, \theta_2 \ldots, \theta_k$.

If $\mu r'$ is the rth moment about origin will be the function of the k parameters $\theta_1, \theta_2 \ldots \theta_k$

Volume : 4 | Issue : 4 | April 2015 • ISSN No 2277 - 8179

and μr' is defined as follows:

$$\mu_r' = \int_{-\infty}^{\infty} x^r f(x; \theta_1, \theta_2 \ldots, \theta_k) dx \text{ , } r = 1, 2 \ldots k.$$

Estimate the parameters θ_1, θ_2...,θ_k in terms of μ_1',μ_2',...,μ_k' by the method of moments consists in solving the k equations μr', r =1,2,...,k and then replacing the population moments μr', r =1,2,...,k by the sample moments mr', r =1,2,...,k, the rth moment about origin in the sample.

$$\text{ie, } \mu_r' = m_r', \; r = 1, 2 \ldots k.$$

$$m_r' = \sum_{i=1}^{n} X \, i^r / n$$

5.2 Graphical Method

Diagrams and graphs are used to present nature of rural taluk size distribution.

Diagrams

Diagrams are based on scale but not confined to points or lines. There are various geometrical shapes such as bars, circles, square, etc. Diagrams are visual presentation of categorical and geographical data. It furnishes only approximate information. Diagrams are more appealing to the eyes and even laymen can understand the concept easily under study.

Graphs

Graphs are more appropriate to represent the time series data and the frequency distribution. Graphs are more precise and accurate than the diagrams. Graphs can be effectively used to study the slopes, rate of change and forecasting. Frequency curve, O give curve, trend line etc., are graphs representing the data related to the study.

6. Techniques

6.1 Statistical measures

Mean

The expected value of a discrete random variable is a weighted average of all possible values of the random variable, where the weights are the probabilities associated with the corresponding values.

The expected value of a discrete random variable **X** with probability mass function f(x) described as follows:

$$E(X) = \sum_x x f(x)$$

Similarly, the expected value of a continuous random variable X with probability density function f(x) described as follows:

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

Variance

The variance of a random variable X, with probability function f(x), is defined as,

$$\text{Where, } E(X^2) = \sum_x x^2 f(x) \text{ , if X is a discrete r.v}$$

$$= \int_{-\infty}^{\infty} x^2 f(x) dx, \text{ if X is a continuous r.v}$$

Coefficient of variation

The coefficient of variation is a percentage and calculated from the average and standard deviation as follows:

$$C.V = \frac{\sqrt{\mu_2}}{\bar{X}} \times 100,$$

$$\text{Where, } \mu_2 = \frac{1}{N} \sum_i f_i (X_i - \bar{X})^2,$$

$$\bar{X} = \frac{\sum_i f_i X_i}{N}$$

Skewness

Skewness is the measure of the deviation of a distribution from symmetry and it is defined as follows:

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} \text{ , Where, } \mu_2 = \frac{1}{N} \sum_i f_i (X_i - \bar{X})^2,$$

$$\mu_3 = \frac{1}{N} \sum_i f_i (X_i - \bar{X})^3$$

Kurtosis

Kurtosis is the measure of the flatness or peakedness of the frequency distribution and it is defined as,

$$\beta_2 = \frac{\mu_4}{\mu_2^2},$$

$$\text{Where, } \mu_2 = \frac{1}{N} \sum_i f_i (X_i - \bar{X})^2,$$

$$\mu_4 = \frac{1}{N} \sum_i f_i (X_i - \bar{X})^4$$

6.2. Test statistics

Chi square test

If $x_1, x_2, \ldots x_n$ are n independent r.v's such that $x_i \sim N(\mu_i, \sigma_i^2)$, for all i=1, 2....n, then the statistic χ^2 is defined as,

$$\chi^2 = \sum \left(\frac{x_i - \mu_i}{\sigma_i}\right)^2 = \sum_{i=1}^{n} Z_i^2 \text{ , where } z_i \sim N(0, 1)$$

It is distributed as a chi-square (χ^2) variate with (n-1) degrees of freedom.

Normal test

A random sample $X^1, X^2, \ldots X^n$ of size n has been drawn from a normal population with mean μ and variance σ^2.The sample mean $\bar{x}$is distributed normally with mean and variance σ^2/n.

ie, $\bar{X} \sim N(\mu, \sigma^2/n)$.The standard normal variate corresponding to the sample mean $\bar{X}$is described as,

$$Z = \frac{\bar{X} - \mu}{\sigma / n}$$

Then the null hypothesis H_0 stating that the sample has been drawn from population with mean and variance σ^2.The standard normal variate Z is distributed as normal distribution with mean 0 and variance 1when H_0 is true. In such a situation the standard normal statistic is called normal test statistic for testing the null hypothesis. When the population variance σ^2 is unknown its estimate (s^2), the sample variance may be used. Similarly, the normal test statistic, **for testing the equality of the two sample means, the normal test statistic is described as follows:**

Let X_1 be the observed taluk rural size and X_2 be the expected taluk rural size, N_1 be the observed no. of rural taluks and N_2 be the expected no. of rural tauks from the population with means $\mu_1$and μ_2 and variances S_1^2 and S_2^2 respectively.

$$Z = \frac{\bar{X}1 - \bar{X}2}{\sqrt{\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}}} \sim N(0, 1)$$

Where, $\overline{X_1}$ = Mean of the observed taluk size distribution.

$\overline{X_2}$ = Mean of the expected taluk size distribution.

Find the Z table value at the desired level of significance and compared with the Z calculated value.

H_0 is accepted if $|Z| < Z_\alpha$

H_0 is rejected if $|Z| > Z_\alpha$

2001 and 2011 census population data have applied to study the nature of taluk size distribution of TAMILNADU state. Rural population in India as well as Tamilnadu state has the growth in nature.

The patterns of the rural taluk size distribution have been described by the four statistical models such as exponential, q-exponential, truncated exponential and lognormal models. These models were fitted the empirical rural taluk size distribution in 2001 and 2011.

6.3. Statistical models

Exponential model

Let X be a random variable having the p.d.f,

Let X be a random variable having the p.d.f,

$$f(x) = \begin{cases} \lambda e^{-\lambda x} ; x>0, \\ 0 \quad ; \text{otherwise} \end{cases}$$

Here, $\lambda>0$ is the parameter of the distribution and is estimated by the method of moments.

The mean and variance are obtained as,

The parameter of the exponential model is,

$$\hat{\lambda} = \frac{1}{\bar{x}}$$

The expression of the corresponding cumulative distribution function is,

$$F_x(x) = P(X \le x) = 1 - e^{-\lambda x}$$

q – Exponential model

Let X be a random variable having the pdf,

$$f(x) = \begin{cases} \frac{a}{q}\left[1 + \frac{q-1}{q} ax\right]^{\frac{q}{q-1}} , x>0 \\ 0, \text{otherwise} \end{cases}$$

Where, a>0 and q>1 are parameters estimated by the method of moments.

The mean and variance are:

$$E(x) = \frac{q}{a(2-q)} , q<2$$

$$V(x) = \frac{q^2}{a^2(2-q)^2(3-2q)} , q < 3/2$$

The parameters of the q-exponential model is estimated using method of moments as,

$$\hat{a} = \frac{2s^2 - s^2}{s(s^2 + s^2)}$$

$$\hat{q} = \frac{2s^2 - s^2}{(2s^2)}$$

The corresponding cumulative distribution of the q-exponential model is,

$$F_x(x) = P(X \le x) = 1 - [1 + \frac{q-1}{q} ax]^{\frac{1}{1-q}}$$

Lognormal model

Let X be a random variable having the lognormal p.d.f,

$$f(x) = \begin{cases} \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}} , x>0 \\ 0; \text{otherwise} \end{cases}$$

The random variable Y=log x is normal random variable having mean μ and variance σ^2 estimated by the method of moments, which log x denotes the natural logarithm of the rural population of taluks.

The estimates of μ and σ^2 are:

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{n} \log_e x_i$$

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n} f(\log_e x_i - \hat{\mu})^2}{N} - \left(\frac{\sum_{i=1}^{n} f \log_e x_i}{N}\right)^2}$$

Where, N is the sample size

The mean and variance of the lognormal model are obtained as,

Volume : 4 | Issue : 4 | April 2015 • ISSN No 2277 - 8179

$$\overline{X} = e^{\mu + \frac{\sigma^2}{2}}$$

$$S^2 = e^{2\mu}\left(e^{2\sigma^2} - e^{\sigma^2}\right)$$

The corresponding cumulative distribution of the lognormal model is,

$$F_X(x) = P(X \le x)$$

$$= \frac{1}{2} + \frac{1}{2}\,\mathrm{erf}\left(\frac{\ln x - \mu}{\sigma\sqrt{2}}\right)$$

Where erf denotes the error function associated with the normal distribution.

Truncated exponential model

Let X be a random variable having the p.d.f of exponential model is given by,

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & ; x>0, \lambda>0 \\ 0 & ; \text{otherwise} \end{cases}$$

The probability density function g(x) of the distribution truncated at X=a is given by,

$$f(x) = \lambda e^{\lambda a} e^{-\lambda x} \quad ; x>a$$

Here, λ is the parameter estimated by the method of moments.

The Mean of the truncated exponential model obtained as

$$\bar{X} = a + \frac{1}{\lambda}$$

The parameter of the truncated exponential model is,

$$\hat{\lambda} = \frac{1}{\bar{x} - a}$$

The expression of the corresponding cumulative distribution function is,

$$F_X(x) = P(X \le x)$$

$$= 1 - e^{\lambda(a-x)}$$

Rural taluk size distribution has studied using 2001 and 2011 TAMILNADU census data. Empirical statistical models such as **exponential, q-exponential, truncated exponential and lognormal** models were proposed to study the nature of distribution of rural taluk size. The empirical results are summarized below:

Empirical Results and Discussion

Rural taluk population data of 2001 and 2011 census are used to describe the rural taluk size distribution and presented in **Table no.1 and Table no.7.**

Rural taluk size distribution – 2001

Table no.1

Rural Taluk Size (in'000)	No. of rural taluks
0-55	7
55-110	39
110-165	64
165-220	36
220-275	21
275-330	22
330-385	7
385	4
Total	**200**

Exponential Model

It is fitted using data in **Table-(1)** as follows:

Parameter of the exponential model are obtained as,

$$\bar{X} = 175.725$$

$$\hat{\lambda} = 0.0057$$

$$f(x) = \begin{cases} (0.0056)e^{(-0.0056)x} & ; x>0 \end{cases}$$

The expected no. of rural taluks and the cumulative distribution function have obtained using the probability density function f(x) and it is presented in **Table no.2** below:

Rural Taluk Size (in'000)	f(x)	Expectedno. of rural taluks	$F_x(x)=P(X)$
0-55	0.2691	54	0.2691
55-110	0.1967	39	0.4658
110-165	0.1437	29	0.6095
165-220	0.1051	21	0.7146
220-275	0.0768	16	0.7914
275-330	0.0561	11	0.8475
330-385	0.0410	8	0.8885
385	0.1115	22	1.0000

Table no.2

Total 1 .0000 200

q – Exponential Model

It is fitted using data in **Table no.1** as follows:

Parameters of the q-exponential model are obtained as,

$$\bar{X} = 175.725$$

$$S^2 = 7780.8294$$

$$\hat{a} = 0.0166 > 0$$

$$\hat{q} = 1.4887 > 1$$

$$f(x) = \begin{cases} \frac{0.0166}{1.4887}\left(1 + \frac{(1.4887-1)}{1.4887} \times 0.0166\right)x)^{\frac{1.4887}{(1-1.4887)}} & ; x > 0 \\ 0; \text{ otherwise} \end{cases}$$

The expected no. of rural taluks and the cumulative distribution function have obtained using the probability density function f(x) and itis

Rural Taluk Size (in'000)	f(x)	Expected no. of rural taluks	$F_x(x) = P(X \leq x)$
0-55	0.4126	83	0.4126
55-110	0.2022	40	0.6148
110-165	0.1136	23	0.7284
165-220	0.0701	14	0.7985
220-275	0.0462	9	0.8447
275-330	0.032	6	0.8767
330-385	0.0231	5	0.8998
> 385	0.1002	20	1.0000

presented in **Table no.3** below:

Table no.3

Total	**1.0000**	**200**

Lognormal Model

It is fitted using data in **Table no. 1** as follows:

The estimates of the parameters μ *and* σ^2 in the normal distribution are obtained as,

$$\hat{\mu} = 5.0321, \hat{\sigma} = 0.5577$$

Based on $\hat{\mu}$ and $\hat{\sigma}$, mean and variance of the lognormal distribution are obtained as,

$$\bar{X} = 179.0383, S^2 = 11696.6330$$

$$f(x) = \begin{cases} \frac{1}{x(0.5577)\sqrt{2\pi}} e^{\frac{-1}{2(0.5577)^2}(\log x - 5.0321)^2} & ; x > 0 \\ 0; \text{ otherwise} \end{cases}$$

The expected no. of rural taluks and the cumulative distribution function have obtained using the probability density function f(x) and it is presented in **Table no.4** below:

Table no.4

Rural Taluk Size (in'000)	f(x)	Expected no. of rural taluks	$F_x(x) = P(X \leq x)$
0-55	0.0331	7	0.0331
55-110	0.2430	49	0.2761
110-165	0.2766	55	0.5527
165-220	0.1889	38	0.7416
220-275	0.1112	22	0.8528
275-330	0.0627	12	0.9155
330-385	0.0352	7	0.9507
> 385	0.0493	10	1.0000
Total	1.0000	200	

Truncated Exponential Model

Rural taluk size distribution -2001

Table no.5

Rural Taluk Size (in'000)	No.of rural taluks
55-110	39
110-165	64
165-220	36
220-275	21
275-330	22
330-385	7
> 385	4
Total	193

It is fitted using data in **Table no.5** as follows:

Parameter of the truncated exponential model are obtained as,

$$\bar{x} = 181.1010$$

$$\hat{\lambda} = 0.0079$$

$$f(x) = \begin{cases} (0.0079)e^{(0.0079)a} e^{(-0.0079)x} & ; x > a \\ 0; \text{ otherwise} \end{cases}$$

Volume : 4 | Issue : 4 | April 2015 • ISSN No 2277 - 8179

The expected no. of rural taluks and the cumulative distribution function have obtained using the probability density function f(x) and itis presented in **Table no.6** below:

Table no.6

Rural Taluk Size (in'000)	f(x)	Expected no. of rural taluks	$F_x(x) = P(X \leq x)$
55-110	0.3524	68	0.3524
110-165	0.2282	44	0.5806
165-220	0.1478	29	0.7284
220-275	0.0957	18	0.8241
275-330	0.6020	12	0.8861
330-385	0.0401	8	0.9262
> 385	0.0738	14	1.0000
Total	1.0000	193	

Rural taluk size distribution -2011

Table no.7

Rural Taluk Size (in'000)	No.of rural taluks
0-55	12
55-110	42
110-165	58
165-220	40
220-275	20
275-330	19
330-385	6
> 385	9
Total	206

Exponential Model

It is fitted using data in **Table no.7** as follows:

Parameter of the exponential model are obtained as,

$$\bar{X} = 174.8786$$

$$\bar{\lambda} = 0.0057$$

$$f(x) = \begin{cases} (0.0057)e^{(-0.0057)x} & ; x > 0 \\ 0; \text{otherwise} \end{cases}$$

Using the probability density function f(x), the expected no. of rural taluks and the cumulative distribution function has obtained and is presented in **Table no.8** below:

Table no.8

Rural Taluk Size (in'000)	f(x)	Expected no. of rural taluks	$F_x(x) = P(X \leq x)$
0-55	0.2691	55	0.2691
55-110	0.1967	40	0.4658
110-165	0.1437	30	0.6095
165-220	0.1051	22	0.7146
220-275	0.0768	16	0.7914
275-330	0.0561	12	0.8475
330-385	0.041	8	0.8885
> 385	0.1115	23	1.0000
Total	1.0000	206	

q – Exponential Model

It is fitted using data in **Table no.7**as follows:

Parameters of the q-exponential model are obtained as,

$$\bar{X} = 174.8786$$

$$S^2 = 89113323$$

$$\bar{\alpha} = 0.0167 > 0$$

$$\bar{q} = 1.4902 > 1$$

$$f(x) = \begin{cases} \frac{0.0167}{1.4902}\left(1 + \frac{(1.4902-1)}{1.4902} \times (0.0167)x\right)^{\frac{1.4902}{(1-1.4902)}} ; & x > 0 \\ 0; \text{otherwise} \end{cases}$$

Using the probability density function f(x), the expected no. of rural taluks and the

cumulative distribution function has obtained and is presented in Table no.9 below:

Table no.9

Rural Taluk Size (in'000)	f(x)	Expected no. of rural taluks	$F_x(x)=P(X \leq x)$
0-55	0.0588	12	0.0588
55-110	0.281	54	0.3178
110-165	0.2515	52	0.5693
165-220	0.1679	34	0.7372
220-275	0.1018	21	0.839
275-330	0.0612	13	0.9003
330-385	0.0567	7	0.9570
> 385	0.0430	13	1.0000
Total	1.0000	206	

Lognormal Model

It is fitted using data in **Table no.7** as follows:

The estimates of the parameters μ and σ^2 in the normal distribution are obtained as,

$$\hat{\mu}=4.9967, \hat{\sigma}=0.6252$$

Based on $\hat{\mu}$ and $\hat{\sigma}$, mean and variance of the lognormal distribution are obtained as,

$$\bar{X}=179.8458$$

$$S^2=15467.4963$$

$$f(x)=\begin{cases}\frac{1}{x(0.6252)\sqrt{2\pi}}\, e^{\frac{-1}{2(0.6252)^2}(\log x-4.9967)^2}; \\ x>0 \\ 0; \text{ otherwise}\end{cases}$$

Using the probability density function f(x), the expected no. of rural taluks and the cumulative distribution function has obtained and is presented in **table no.10** below:

Table no.10

Rural Taluk Size (in'000)	f(x)	Expected no. of rural taluks	$F_x(x)=P(X \leq x)$
0-55	0.4167	86	0.4167
55-110	0.2023	42	0.619
110-165	0.1131	23	0.7321
165-220	0.0695	14	0.8016
220-275	0.0457	9	0.8473
275-330	0.0316	7	0.8789
330-385	0.0227	5	0.9016
> 385	0.0984	20	1.0000
Total	1.0000	206	

Truncated Exponential Model

Rural taluk size distribution -2011

Table no.11

Rural Taluk Size (in'000)	No. of rural taluks
55-110	42
110-165	58
165-220	40
220-275	20
275-330	19
330-385	6
> 385	9
Total	194

It is fitted using data in **Table no.11** as follows:

Parameter of the truncated exponential model are obtained as,

$$\bar{x}=183.9948, \quad \hat{\lambda}=0.0077$$

$$f(x)=\begin{cases}(0.0077)e^{(0.0077)a}\, e^{(-0.0077)x}; & x>a \\ 0; \text{ otherwise}\end{cases}$$

Using the probability density function f(x), the expected no. of rural taluks and the cumulative distribution function has obtained and is presented in **Table no.12** below:

Volume : 4 | Issue : 4 | April 2015 • ISSN No 2277 - 8179

Table no.12

Rural Taluk Size (in'000)	f(x)	Expected no.of rural taluks	$F_x(x)$= P(X≤ x)
55-110	0.3452	67	0.3452
110-165	0.2260	44	0.5712
165-220	0.1480	29	0.7192
220-275	0.0969	19	0.8161
275-330	0.0634	12	0.8795
330-385	0.0415	8	0.9210
> 385	0.0790	15	1.0000
Total	1.0000	194	

Model wise estimates of the parameters are presented in Table no.15 for the Year 2001 and 2011

Table no.13

[Rural taluk population]

It is fitted using data in **Table no.11** as follows:

Parameter of the truncated exponential model are obtained as,

$$\bar{x} = 183.9948, \quad \hat{\lambda} = 0.0077$$

$$f(x) = \begin{cases} (0.0077)e^{(0.0077)a}\, e^{(-0.0077)x} & ; x > a \\ 0; \text{ otherwise} \end{cases}$$

Using the probability density function f(x), the expected no. of rural taluks and the cumulative distribution function has obtained and is presented in **Table no.12** below:

Table no.12

Rural Taluk Size (in'000)	f(x)	Expected no.of rural taluks	$F_x(x)$= P(X≤ x)
55-110	0.3452	67	0.3452
110-165	0.2260	44	0.5712
165-220	0.1480	29	0.7192
220-275	0.0969	19	0.8161
275-330	0.0634	12	0.8795
330-385	0.0415	8	0.9210
> 385	0.0790	15	1.0000
Total	1.0000	194	

r	Expo nential	q-Exponential		Lognormal		Truncated Expo nential
	$\bar{\lambda}$	$\hat{\alpha}$	$\hat{q}$	$\hat{\mu}$	$\hat{\sigma}$	$\bar{\lambda}$
l	0.0057	0.0166	1.4887	5.0321	0.5577	0.0079
l	0.0057	0.0167	1.4902	4.9967	0.6252	0.0077

Inference based on statistical tests

1. Normal test

Table no.14

Model	Statistical Z value		Inference at 5% level of significance	
	Year 2001	Year 2011	Year 2001	Year 2011
Exponential	1.56	1.2753	Accepted at 5% level	Accepted at 5% level
q-exponential	5.049	4.419	Rejected at 5% level	Rejected at 5% level
Lognormal	0.1810	0.2775	Accepted at 5% level	Accepted at 5% level
Truncated exponential	0.5407	0.608	Accepted at 5% level	Accepted at 5% level

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}} \sim N(0,1)$$

H_0: $\mu_1 = \mu_2 = 0$. ie., the samples come from the same population.

The results show that the lognormal model is preferred to represent the rural taluk size distribution than the other models in the year 2001 and 2011 because the normal test values are less than the expected value at 5% level of significance.

2. Chi-square test of goodness of fit

Table no.15

Model	Chi-square value		Inference at certain level of significance	
	Year 2001	Year 2011	Year 2001	Year 2011
Exponential	121.2778	88.6838	Rejected at 5% level	Rejected at 5% level
q-exponential	249.5404	205.4868	Rejected at 5% level	Rejected at 5% level
Lognormal	15.5972	8.6082	Accepted at 0.5% level	Accepted at 5% level
Truncated exponential	39.2492	24.9911	Rejected at 5% level	Rejected at 5% level

$\chi^2 = \sum_{i=1}^{n} [(O_i - E_i)^2 / E_i]$. When $\sum_{i=1}^{n} Oi = \sum_{i=1}^{n} Ei$

H0: The fit of the exponential, q-exponential, lognormal & truncated exponential models are good fit to the rural taluk size distribution.

The results show that the lognormal model is preferred to rep-

resent the rural taluk size distribution than the other models in the year 2001 and 2011 because the test values are less than the expected value at 0. 5% level of significance in 2001 and 5% level of significance in 2011.

The information given in **Table no. 1 to Table no.6** is used to draw the following **Fig.1.**

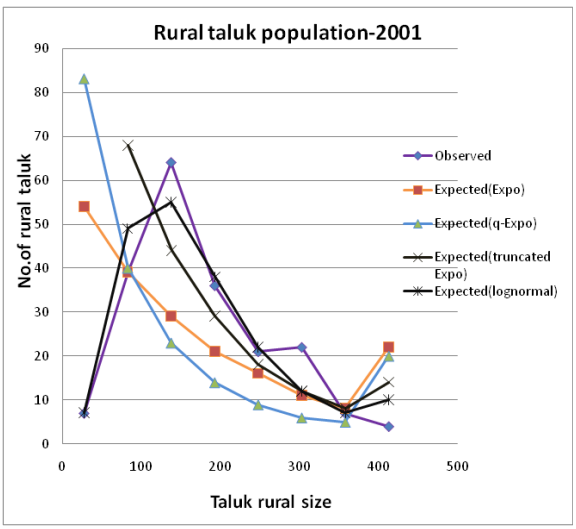

Fig.1

It indicates that the expected curve of lognormal model is very closed to the observed curve but the expected curve of truncated exponential, exponential and q-exponential model are deviated from the observed curve. Thus, lognormal model is more suitable than the other models and also it is the best description of the growth nature of the rural taluk population in the year 2001.

The information given in **Table no. 8 to Table no. 13** are used to draw the following **Fig.2.**

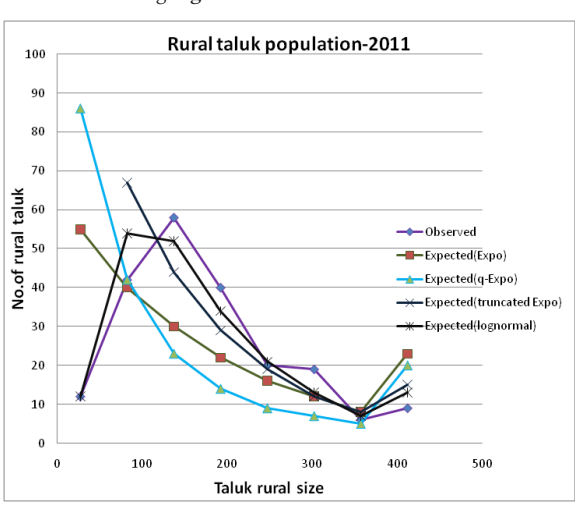

Fig.2

It indicates that the expected curve of lognormal model is very closed to the observed curve but the expected curve of truncated exponential, exponential and q-exponential model are deviated from the observed curve. Thus, lognormal model is more suitable than the other models and also it is the best description of the growth nature of the rural taluk population in the year 2011.

Summary and conclusion

Summary

Background papers deals with the city size distribution. To study the nature of rural taluk size distribution empirically, 2001 and 2011 Tamilnadu census taluk size population data have been used. Rural taluk size distributions are all skew in nature. Hence skew models such as exponential, q-exponential, truncated exponential and lognormal models were proposed to carry out the empirical study. The empirical results based on these four models have been presented as follows:

The average rural taluk size were obtained and presented as follows:

Models	Average rural taluk size	Empirical and fitted models		
			Difference	
	2001	2011	2001	2011
Empirical model	175.725	174.8786	-	-
Exponential	159.775	160.7281	15.950	14.150
q-exponential	127.600	126.5534	48.125	48.325
lognormal	174.075	172.2087	1.650	2.670
Truncated exponential	175.9715	178.0412	-0.2465	-3.1626

The average lognormal taluk size distribution and truncated exponential taluk size distribution are deviated in least from the average taluk size of empirical distribution in both the years. Other models having larger deviations in both the years. It indicate that lognormal and truncated exponential models may be preferred than the other models.

The variability of taluk size distribution fitted statistical models such as exponential, q-exponential, truncated exponential and lognormal models are measured in terms of coefficient of variation.

Models	Coefficient of variation	
	2001	2011
Exponential	80%	80%
q-exponential	80%	99%
lognormal	54%	58%
Truncated exponential	58%	58%

The coefficient of variation in both the years due to lognormal and truncated exponential model are minimum than the exponential and q-exponential models. It shows that the lognormal and truncated exponential model may be preferred than the other models because less amount of variation in lognormal taluk size distribution and truncated exponential taluk size distribution have been observed.

The skew nature of the models may be studied in terms of measures of skewness. β_1 It is obtained and presented as follows.

Models	Skewness measure(β_1)	
	2001	2011
Exponential	0.5703	0.5549
q-exponential	5.3046	1.5159
lognormal	0.8219	1.4452
Truncated exponential	1.0275	0.9727

β_1 Values are greater than zero in all the models. Hence all the models are referred as a skew model. Hence these models may be used to describe the taluk size distribution.

The kurtosis of the four models are obtained and presented as follows:

Volume : 4 | Issue : 4 | April 2015 • ISSN No 2277 - 8179

Models	Kurtosis measure(β_2)	
	2001	2011
Exponential	2.3324	2.3144
q-exponential	7.5021	3.2446
lognormal	3.2571	3.0492
Truncated exponential	2.965	2.8928

Lognormal and truncated exponential model having values very closed to three than the other models in both the years 2001 and 2011.It indicate that these two models are moving towards the symmetric nature in the long run.

The inference based on the normal test statistic indicated that lognormal and truncated exponential rural taluk size distribution may be preferred to describe the rural taluk size distribution in both the years because the null hypothesis in the normal test have been accepted at 5% level of significance. The acceptance of null hypothesis indicates that the differences between the average taluk sizes in the two models are nil.

The test for goodness of fit of the models exponential, q-exponential, truncated exponential and lognormal models have been tested using statistic and inferred that the fitness of lognormal model is inferred as a good fit for the rural taluk size distribution in both the years than the other models.

Fig.1 and Fig.2 indicate that the taluk size distribution represents the lognormal model have been observed as very closed to the empirical taluk size distribution than the other models.

Conclusion

In the nutshell lognormal model has been identified as a best model for representing the rural taluk size distribution because

* having average taluk size very closed to the empirical average taluk size.
* **having least amount of variation in taluk size than the other models.**
* **having the skew characteristics such as** β_1 values greater than zero.
* **moving towards symmetric nature in the long run because** β_1 values are very closed to three.
* Goodness of fit is good at 5% level of significance.
* Graphically its curve having very closed to the curve of the empirical taluk size distribution.

Thus, the lognormal model may be used to describe the nature of the rural taluk size distribution. The rural taluk size population may be analyzed using multivariate models for further study.

REFERENCE

1. DIRECTORATE OF CENSUS OPERATIONS, Tamilnadu, Chennai: primary census Abstract, Census-2001&2011. | 2. Gonzalez-Val, Rafael and Ramos, Arturo and Sanz, Fernando and Vera-Cabello, Maria (2013): Size Distributions for All Cities: Which One is best? JEL: C13, C16, R00.Regional science. MPRA_paper_45019 | 3. R.K .Gupta (2004).RURAL DEVOLOPMENT IN INDIA. Atlantic publishers and distributors, New Delhi. | 4. S.C.Gupta and V.K. Kapoor (2007). FUNDAMENTALS OF MATHEMATICAL STATISTICS. Sultan and sons, New Delhi. | 5. Keith Briggs (2006)."Modeling train delays with q-exponential functions". Physica A: Statistical Mechanics and its Applications.Volume 378, Issue 2, 15 May 2007, Pages 498–504.Elsevier. | 6. Malacarne, L. C., R. S. Mendes and E. K. Lenzi (2001). "q-exponential distribution in urban agglomeration," Physical Review E 65(017106), 1–3. | 7. V.K. Rohatgi (1976). AN INTRODUCTION TO PROBABILITY THEOREY AND MATHEMATICAL STATISTICS. | 8. H.C.Saxena, P.U.Surendran (1990).STATISTICAL INFERENCE. S.cand and company, Ltd., New Delhi. | 9. S.Suddendu Biswas, (2011).MATHEMATICAL STATISTICS .Narosa publishing house, New Delhi. | **Website references** | 1. http://en.wikipedia.org/wiki/Rural_development. | 2. www.tnrd.gov.in/policynotes/rd-policynotes-2011-2012. | 3. http://en.wikipedia.org/wiki/List_of_Developmental_administrative_units_of_Tamil_ Nadu. | 4. www.businessdictionary.com/definition/rural.html. | 5. www.brainpopjr.com/socialstudies/communities/ruralsuburbanandurban | 6. www.answers.com › ... › Personal Finance › Money Management › Banking | 7. www.definitions.net/definition/RURAL | 8. censusindia.gov.in/.../paper2/data files/India/Rural_Urban_2011.pdf | 9. www.variation.com/cpa/help/hs114.htm | |