International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2014): 5.611

Poisson Regression Model of Administrative City Size

M. Sakthi

Assistant Professor, Periyar Government Arts College, Cuddalore, Tamilnadu, India

Abstract: ***Urbanization is a process of growth of the region in both ways namely size and area. City is an urban area having significant role in the growth of the urbanization as well as city sizes. Administrative city size is the function of its age and area and other factors such as education, transport facilities. Poisson regression model is proposed to study the growth of the administrative city sizes.***

Keywords: Administrative city, Poisson regression, overdispersion, wald test, likelihood ratio test

1. Introduction

City is a relatively large and permanent settlement. Cities generally have complex systems for sanitation, utilities, land usage, housing, and transportation. The concentration of development greatly facilitates interaction between people and businesses, benefiting both parties in the process. Administrative cities were the habitations of the state rulers. Their major cultural role was to serve as the locus of state administration. Administrative cities also had a qualitatively different demographic and social complexity. Such cities were nodes of communication and transportation and centres of commerce, crafts, and other economic functions for the surrounding countryside. India's urban population is growing at an average rate of around 3 percent per annum. India Census(2001), it has almost doubled during the period between 1981 and 2001 from 160 million to 285 million. In terms of percentage of total population, the urban population has gone up from 17% in 1951 to 29% in 2001 and is expected to increase up to around 37% by the year 2021. Consequently, the number and size of cities have also increased considerably. According to 2001 India census, the extent of the Tamilnadu State is 130,058 sq.km of which the urban area accounts for 12,525 sq.km. (India Census 2001) Tamilnadu is the most urbanised state in India, and its urban population has grown from 12.46 million in 1971 to 27.48 million in 2001. City size distribution is incorporated with economic development and cultural changes(Brain J.L. Berry 1961). The size distribution of cities is a function of the age distribution of cities (Richardson H.W 1973). New testing procedure for testing the validity of the rank size rule for city size distribution (Gershon Alperovich 1984, 1989). Performance of Poisson and negative binomial (NB) regression models used the relationship between truck accidents and geometric design of road sections (S.P Miaou1994). Regression model based on the Poisson distribution as a tool for resolving common problems in analyzing aggregate crime rates (D. Wayne Osgood 2000). Age of the city has not been taken to describe the city size distribution. Poisson regression model have not been applied to study the city sizes. Age of the city is incorporated to describe the city size distribution. Poisson regression describe the city sizes based on the factors age of the city, area of the city, education and transport of the city.

2. Objective

General Objective

To propose the regression model for describing the administrative its size which is the function of age of the city, area of the city, education and transport facilities.

Specific Objective

- To analyse the growth of administrative city size empirically using Poisson regression model.
- To test the suitability of the regression models for the administrative city population data in Tamil Nadu state.

3. Model

A model is a simplification of reality. A model is an abstract representation of reality which clears what is relevant to a particular question at a particular time and neglects all other aspects. 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.

Poisson Regression Model

The standard model for count data is the Poisson regression model, which is a nonlinear regression model. This regression model is derived from the Poisson distribution by allowing the intensity parameter μ to depend on covariates (regressors). A regression model based on the Poisson distribution follows by conditioning the distribution of y_i on a *k*-dimensional Vector of covariates, $X_i = [x_{1i}, \dots, x_{ki}]$ and parameters β, through a continuous function $\mu(x_i, \beta)$, such that $E[y_i \mid X_i] = \mu(x_i, \beta)$

That is, y_i given X_i is Poisson-distributed with density

$$f(y_i \mid X_i) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!} \qquad y_i = 0,1,2,\dots$$

In the log-linear version of the model the mean parameter is parameterized as

$$\mu_i = \exp(X'\beta_i) \quad (3)$$

to ensure $\mu > 0$. Equations (2) and (3) jointly define the Poisson (loglinear) regression model.

Testing For Overdispersion in Poisson regression

Deviance and Pearson Chi-Square divided by the degrees of freedom are used to detect overdispersion or underdispersion in the Poisson regression. Values greater than 1 indicate overdispersion, that is, the true variance is bigger than the mean, values smaller than 1 indicate underdispersion, the true variance is smaller than the mean. Evidence of underdispersion or overdispersion indicates inadequate fit of the Poisson model. We can test for overdispersion with a likelihood ratio test based on Poisson and negative binomial regressions. This tests equality of the mean and the variance imposed by the Poisson distribution against the alternative that the variance exceeds the mean.

Goodness of Fit Test statistics

The Pearson statistic $P_p = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i}$ and the deviance, or G, statistic

$$D_p = \sum_{i=1}^{n} \left\{ y_i \ln\left(\frac{y_i}{\hat{\mu}_i}\right) - (y_i - \hat{\mu}_i) \right\}$$

approximately distributed as chi-square distribution with *n - k* degrees of freedom. The Pearson statistic is often used as a test statistics for testing overdispersion. The null hypothesis is to be considered that the model is fit well for the data.

The likelihood ratio test statistics

The Likelihood Ratio test is performed by estimating two models and comparing the fit of one model to the fit of the other. The formula for the Likelihood Ratio test statistic is described as

$$LR = -2\ln\left(\frac{\text{likelihood for null model}}{\text{likelihood for alternative model}}\right)$$
$$= -2\ln(\text{likelihood for null model}) + 2\ln(\text{likelihood for alternative model})$$

The resulting test statistic is distributed as chi-square distribution, with degrees of freedom equal to the number of parameters that are constrained. Consider the null hypothesis is that all of the regression coefficients are simultaneously equal to zero

Wald Test:

The Wald test is the testing of significance of particular explanatory variables in a statistical model. If for a particular explanatory variable, or group of explanatory variables, the Wald test is significant, then we would conclude that the parameters associated with these variables are not zero, so that the variables should be included in the model. If the Wald test is not significant then these explanatory variables can be omitted from the model. Under the Wald statistical test, the maximum likelihood estimate $\hat{\theta}$ of the parameter(s) of interest θ is compared with the proposed value θ_0, with the assumption that the difference between the $\hat{\theta}$ and θ_0 will be approximately normal. Typically the square of the difference is compared to a chi-squared distribution. In the univariate case, the Wald statistic is $\sum \frac{(\hat{\theta} - \theta_0)^2}{\text{var}(\hat{\theta})}$. Alternatively, the difference can be compared to a normal distribution. In this case the test statistic is $\sum \frac{\hat{\theta} - \theta_0}{SE(\hat{\theta})}$ where $SE(\hat{\theta})$ is the standard error of the maximum likelihood estimate.

Data Analysis

Source of data

Data about the administrative cities population, age of cities, area of cities, education and transport are collected from Tamil Nadu census report for the year 1971, 1981, 1991 and 2001. Empirical analysis are carried out using the information given in the table-1 and presented the results as follows

Table 1: Deviance and Pearson chi-square value of the Poisson regression

Year	Pearson chi-square value	Deviance value	Expected chi-square value at 1% level of significance	Inference
1971	9.537	11.07	24.725	Accepted H_0
1981	12.775	14.356	27.6883	Accepted H_0
1991	21.601	23.927	34.8053	Accepted H_0
2001	45.704	44.941	46.963	Accepted H_0

Table No.1 gives the result of deviance and Pearson chi-square value. The value of Pearson chi-square and deviance are compared with the chi-square table value with their corresponding degrees of freedom. The null hypothesis is accepted. Hence there is no overdispersion. So the Poisson regression is the appropriate model for the data.

Table 2: Likelihood ratio test value of the Poisson regression

Year	LR statistic Value	Expected LR statistic value at 1% level of significance	Inference
1971	56.517	9.21034	Rejected H_0
1981	68.52	9.21034	Rejected H_0
1991	108.497	9.21034	Rejected H_0
2001	69.222	9.21034	Rejected H_0

Table No.2 yields the LR test value which is significant with 4 df. Hence at least one predictor variable regression coefficient is not equal to zero in the model.

Table 3: Wald chi-square value of the parameters of the Poisson regression

Parameter	*1971*	*1981*	*1991*	*2001*	*Expected chi-square value at 1% level of significance*	*Inference*
Intercept	86.654	192.652	304.987	399.499	6.6349	Rejected H_0
Age	17.483	33.592	63.561	54.419	6.6349	Rejected H_0
Area	41.338	34.213	37.571	28.044	6.6349	Rejected H_0

Table No.3 gives the Wald test of the Poisson regression model, which indicates that the slope of the coefficients is significantly different from zero. So the variables age, area should be in the model and education, transport are not much influenced for the growth of the administrative city size.

4. Summary and Conclusion

Various literatures about the city size distribution have been reviewed and observed these age of the city, area of the city are the main determinants of city size. The Poisson regression model have been used to describe administrative city sizes. Likelihood ratio test, Overdispersion tests such as Pearson chi-square test and Deviance test were used to test the goodness of fitting the model. Wald test has been used to find whether the individual predictor variables are in the model or not.

The Poisson regression model has applied empirically to study the growth of the administrative city population. It has been inferred that the Poisson regression showed the growth of the city which is the function of the age of the city and area of the city.

References

[1] Gershon Alperovich(1984), The size distribution of cities: On the empirical validity of the rank size rule, Journal of Urban Economics, vol. 16, 232 – 239.
[2] Gershon Alperovich(1989), The distribution of City size: A sensitivity analysis, Journal of Urban Economics, vol. 25, pp. 93 – 102.
[3] Brain J.L. Berry(1961), City size distribution and economic development, Economic Development and cultural change, vol. 9, N0. 4, 573 – 587
[4] A. Colin Cameron and Pravin K. Trivedi (1998) “Regression analysis of count data” Cambridge university press, NewYork.
[5] C. B. Dean(1992), Testing for Overdispersion in Poisson and Binomial regression models, Journal of the American statistical association, vol. 87, No. 418, pp. 451 – 457.
[6] C. Dean and J. F. Lawless(1989), Tests for detecting Overdispersion in Poisson regression models, Journal of the American statistical association, vol. 84, No. 406, pp. 467 – 472.
[7] S. Gurmu(1971), Testing for detecting overdispersion in the positive Poisson regression model, journal of Business & Economic statistics, vol. 9, No. 2, pp. 215-222.
[8] Joseph M. Hilbe and William H. Greene(2008), Count response regression models, Handbook of Statistics, vol. 27, pp. 210 – 252.
[9] S.P.Miaou(1994), The relationship between truck accdidents and geometric design of road sections: Poisson versus negative binomial regressions, Accid. Anal. Prev., vol. 26, No.4, pp. 471-482.
[10] Norman Breslow(1990), Tests of hypothesis in Overdispersed Poisson regression and other quasi-likelihood models, vol. 85, No. 410, pp. 565 – 571.
[11] Harry W. Richardson(1973), Theory of the distribution of city sizes: Review and prospects, Regional studies, vol. 7, pp. 239 – 251.
[12] Wayne Osgood(2000), Poisson-Based regression analysis of aggregate crime rates, journal of quantitative criminology, vol. 16, No.1, pp. 21 – 43.