Available online at http://www.journalcra.com

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

International Journal of Current Research
Vol.9, pp.058-061, October, 2010

ISSN: 0975-833X

RESEARCH ARTICLE

A STOCHASTIC MODEL FOR THE PREDICTION OF MANPOWER USING INCOMPLETE MANPOWER DATA IN TAMILNADU SOFTWARE INDUSTRY

Susiganeshkumar, E. and Elangovan, R.

Department of Statistics, Annamalai University, Annamalai Nagar – 608 002

ARTICLE INFO

Article History:
Received 13th July, 2010
Received in revised form
17th August, 2010
Accepted 23rd September, 2010
Published online 9th October, 2010

Key words:
Completed Length of Service
Manpower planning
Regarding pay
Perquisites
Promotion and targets

ABSTRACT

In any organization it usually happens that whenever the policy decisions regarding pay, perquisites, promotion and targets of work or sales to be achieved are revised, then, there will be exit of personnel, which in other words is called the wastage. In manpower planning, one of the most important variables is Completed Length of Service (CLS) on leaving a job, since it enables us to predict staff turnover. The data on CLS are often incomplete due to left truncation as well as right censoring. Right censoring occurs when a number of people have not yet left when data collection is terminated. Left truncation arises when some people are already in service at the commencement of data collection. For such data much of work has been done on both non-parametric and parametric estimation. In this paper, a Stochastic model for prediction of manpower using incomplete data in Tamilnadu software industry has been discussed.

INTRODUCTION

In manpower planning, we are concerned with the description and prediction of the behavior of large groups of people. Such data is well suited to a mathematical modeling approach since, although individual behavior is unpredictable, when aggregated, the data is seen to follow statistical patterns which may readily be quantified. For a detailed study, refer to McClean (1991).

In manpower planning, one of the most important variables is completed length of service on leaving a job, since it enables us to predict staff turnover. The data on CLS are often incomplete due to left truncation as well as right censoring. Right censoring occurs when a number of people have not yet left when data collection is terminated. Left truncation arises when some people are already in service at the commencement of data collection. For a detailed study refer to McClean and Gribbin (1987). The Tamilnadu Software Industry has undergone a substantial expansion over the past few years. Historically Tamilnadu has been in an enviable position regarding the quality and quantity of manpower in its software industry. In this paper, data were obtained from 300 software companies in Tamilnadu, 150 were IT Users, 75 were IT Training Providers and 75 were IT Employment Agencies, and the data gives the details about the CLS of the individuals. The data were collected between 24th March 2000, say t_0 (commencement of the period), and 21st August 2006 say t_1 (end of the period). A more detailed discussion for the growth of manpower for the Tamilnadu Software Industry, flows in a manpower system with a special reference to Tamilnadu

***Corresponding author:** *susijk@rediffmail.com*

Software Industry, refer to Susiganeshkumar and Elangovan (2009a), Susiganeshkumar and Elangovan (2009b). In this paper a Stochastic model for prediction of manpower using incomplete data in Tamilnadu software industry has been discussed. Estimation for incomplete manpower data, can also be seen in Kaplan and Meier (1958), Kalbfleisch and Prentice (1980). For arbitrarily grouped, censored and truncated data refer to Turnbull (1976).

Notations

1. $t_1, t_2, t_3, \ldots, t_n$ are the n observed completed length of service times.
2. $S_1, S_2, S_3, \ldots, S_g$ are the g observed right censored observations.
3. $r_1, r_2, r_3, \ldots, r_m$ are the m observed completed length of service which are left truncated at $l_1, l_2, l_3, \ldots, l_m$.
4. $q_1, q_2, q_3, \ldots, q_k$ are the k right censored observation which are left truncated at $p_1, p_2, p_3, \ldots, p_k$.
5. f(.) is the p.d.f of failure times.
6. F(.) is the survival function of failure times.

RESULTS

The two-parameter Gamma distribution has p.d.f.

$$\frac{1}{\alpha\,\Gamma(\mathrm{k})}\left(\frac{1}{\alpha}\right)^{k-1}\exp\left(\frac{-t}{\alpha}\right) \quad ;t>0 \qquad \ldots (3.1)$$

where α>0 and k>0 are unknown scale and shape parameters, respectively.

The survivor function for the Gamma distribution in eqn. (3.1) is

$$\mathrm{S(t)} = \mathrm{Q}\left(k, \frac{1}{\alpha}\right)$$

where

$$Q(k,x) = 1 - I(\mathrm{k},\mathrm{x}) = \frac{1}{\Gamma(\mathrm{k})}\int_x^{\infty} u^{k-1}\, e^{-u}\, du \;..\,(3.2)$$

The appropriate likelihood function in the continuous case

$$\mathrm{L} = \prod_{\mathrm{i}=1}^{\mathrm{n}} f\left(t_i\right)^{\delta_i} S\left(L_i\right)^{1-\delta_i} \qquad \ldots (3.3)$$

Where

t_i is a life time.

L_i is a censoring time (both right censored and left truncated).

$$\delta_{\mathrm{i}} = \begin{cases} 1, \text{if individual i's lifetime is observed} \\ 0, \text{if it is censored} \end{cases}$$

Using this result given in eqn. (3.3). The likelihood function for the Right Censored (RC) and Left Truncated (LT) observations is then,

$$L(k,\alpha) = \left[\prod_{i=D}\frac{1}{\alpha\Gamma(k)}\left(\frac{t_i}{\alpha}\right)^{k-1} e^{-t_i/\alpha}\right]\left[\prod_{i\in C} Q\left(k, \frac{t_i}{\alpha^*}\right)\right] \ldots\ldots (3.4)$$

Letting

$\bar{\mathrm{t}} = \sum_{\mathrm{i}\in \mathrm{D}} \frac{\mathrm{t_i}}{\mathrm{r}}$ and $\tilde{\mathrm{t}} = \left(\prod_{i\in D} t_i\right)^{1/r}$ be the arithmetic means and geometric means of the observed data. The log likelihood function is written as,

$$\log\mathrm{L}(\mathrm{k},\alpha) = -\mathrm{rk}\log\alpha - \mathrm{r}\log\Gamma(\mathrm{k}) + \mathrm{r}(\mathrm{k}-1)\log\tilde{\mathrm{t}} - \frac{\mathrm{r}\bar{\mathrm{t}}}{\alpha} + \sum_{i\in C}\log\left[Q\left(k, \frac{t_i}{\alpha}\right)\right] \;..\,(3.5)$$

Since the derivatives of log L are tedious to calculate, it is often simplest to maximize log L directly by a method that does not require formulas for derivatives. Incomplete Gamma integrals must be calculated using computer routines by SAS program, and also using a detailed procedure suggested by Lowless (1982). The likelihood function also was maximized using a Quasi-Newton Method as suggested by McClean and Gribbin (1987).

A number of parametric and non-parametric models have been fitted the manpower duration data, refer to Silcock (1954). Fitting of Gamma distribution is very difficult since it involves incomplete gamma integral and so the data is fitted for the particular case of exponential distribution. However the formulae are easy to compute and we may, therefore, use a goodness of fit test to compare other distributions with the Exponential. For the Exponential distribution based on the notations discussed in section 2 and procedure suggested by McClean and Gribbin (1987), the likelihood function is therefore,

$$L = \left(\prod_{i=1}^{n} f\left(t_i\right)\right)\left(\prod_{i=1}^{g} F\left(S_i\right)\right)\left(\prod_{i=1}^{m}\frac{Fq_i}{F(l_i)}\right)\left(\prod_{i=1}^{k}\frac{Fq_i}{F(p_i)}\right)$$

Where
$F(x) = e^{-\lambda x}$ and $f(x) = \lambda e^{-\lambda x}$.

$$L=\left(\lambda^{n} e^{-\lambda \sum t_i}\right)\left(e^{-\lambda \sum S_i}\right)\left(\lambda^{m} e^{-\lambda \sum (r_i - l_i)}\right)\left(e^{-\lambda \sum (q_i - p_i)}\right)$$

Differentiating gives the maximum likelihood,

$$\hat{\lambda} = \frac{n+m}{\sum t_i + \sum S_i + \sum (r_i - l_i) + \sum (q_i - p_i)}$$

$$= \frac{\text{Total leavers}}{\text{Total observed service times}}$$

The most widely used distribution for CLS until leaving are the mixed exponential distribution suggested by McClean (1976), Bartholomew (1982) and the log normal distribution suggested by Lane and Andrew (1955), Bartholomew and Forbes (1979). For the collected data from the Tamilnadu Software Industry, the results were shown in the following Table 1.

12.3 years since the commencement of the Tamilnadu Software Industry namely the origin. It is seen from the data set that

$$\sum_{i=1}^{m=17} \log r_i = 0.02, \quad \sum_{i=1}^{n=19} \log t_i = 2.12, \quad \sum_{i=1}^{k=236} \log q_i = 2.88 \text{ and } \sum_{i=1}^{g=304} \log S_i = 3.50.$$

$\hat{\lambda}$ *is* estimated as 0.0352.

THE GOODNESS OF FIT TEST

To test whether our observed data could have come from a particular completed length of service distribution with distribution function F (.). This distribution function is fitted to data collected between times t_0 and t_1 and the model is then used to predict how many of the staff present at t_1 will still be there at a later time t_2. The goodness of fit is tested by comparing this prediction with the

Table 1. CLS (in years) of the individuals belonging to different categories of Tamilnadu Software Industry

Case (i) *Complete data (t_i)*	8.6, 4.3, 7.4, 8.3, 4.6, 5.7, 3.1, 11.8, 2.7, 4.8, 6.10, 8, 7.9, 5.6, 2.5, 3.8, 10.6, 12.3, 13.4
Case (ii) Left truncated (r_i)	10.7, 11.2, 8.7, 6.4, 5.8, 3.2, 2.6, 2.9, 4.7, 6.1, 7.4, 6.3, 2.4, 3.7, 5.4, 6.8, 11.3
Case (iii) Right censored (S_i)	12.1, 10.2, 13.2, 13.2, 13.2, 13.2, 11.2, 11.3, 9, 8, 10.7, 15.8, 15.8, 15.8, 15.8, 15.8, 13.8, 12, 15.8, 15.8, 14.2, 13.2, 13.8, 11, 13, 12.9, 10.0, 12.5, 15.8, 15.8, 15.8, 15.8, 11, 14.9, 12.0, 13.4, 15.8, 15.8, 15.8, 15.8, 15.8, 6, 8, 11.4, 11.4, 11.4, 11.4, 11.4, 11.4, 13.2, 14.3, 15.0, 15.0, 12.3, 12.0, 12.1, 12.1, 12.1, 9.3, 9.3, 3.2, 3.2, 4.3, 4.5, 3.2, 3.2, 8.7, 12.3, 15.8, 15.8, 15.8, 15.8, 15.8, 15.8, 15.8, 15.8, 15.8, 15.8, 15.8, 15.8, 11.9, 10, 11.4, 11.4, 11.4, 11.4, 11.4, 12.3, 12.3, 12.3, 12.3, 12.3, 11.9, 11.8, 11.7,11.7, 11.4, 11.4, 11.4, 11.4, 11.3, 11.5, 11.5, 9.6, 9.6, 9.6, 9.6, 9.6, 9.6, 9.6, 4.1, 9.4, 8.7, 8.7, 8.7, 8.7, 8.7, 8.7, 8.7, 10.1, 10.1, 8.8, 8.8, 8.8, 8.8, 8.8, 8.8, 8.8, 8.8, 3.6, 3.6, 3.6, 3.6, 3.6, 3.6, 3.6, 12.3, 12.4, 12.3, 12.4, 12.3, 12.3, 12.3, 12.3, 12.3, 12.3, 12.3, 8.7, 8.7, 8.7, 8.7, 8.7, 8.7, 8.7, 8.7, 8.7, 8.7, 6.4, 6.4, 6.4, 6.4, 6.4, 6.4, 6.4, 6.4, 7.5, 7.5, 7.5, 7.5, 7.5, 6, 11.8, 11.8, 11.8, 11.8, 11.8, 11.8, 11.8, 11.8, 11.8, 11.8, 11.8, 15.8, 7.5, 7.6, 7.6, 7.6, 7.6, 7.6, 7.6, 7.6, 15.8, 15.8, 15. 8, 15.8, 15.8, 15.8, 15.8, 15.8, 15.8, 15.8, 15.8, 15.8, 15.8, 4.3, 4.3, 4.3, 5.2, 5.2, 3.2, 3.2, 7.5, 7.5, 7.5, 7.5, 7.5, 6.3, 6.3, 6.3, 6.3, 6.3, 6.3, 6.3, 12.2, 12.1, 12.2, 12.2, 12.2, 12.2, 12.2, 12.2, 12.2, 12.2, 12.2, 15.8, 15.8, 15.8, 15.8, 15.8, 15.8, 15.8, 15.8, 15.8, 15.8, 15.8, 15.8, 15.8, 15.8, 10.2, 10.3, 10.2, 10.2, 10.2, 10.2, 10.2, 9.6, 9.6, 9.6, 9.6, 9.6, 9.6, 9.6, 9.6, 3.7, 3.7, 3.7, 3.7, 3.7, 3.7, 3.7, 9.1, 7.6, 7.5, 7.5, 7.5, 7.5, 7.5, 7.1, 6.3, 6.3, 6.3, 6.3, 6.3, 6.3, 15.8, 15.8, 15.8, 15.8, 15.8, 15.8, 15.8, 15.8, 2.4, 2.4, 2.4, 2.4, 2.4, 2.4, 2.4, 4.6, 6, 4.6, 4.6, 1.2, 2.3
Case (iv) Left truncated and Right censored (q_i)	There are 236 individuals who belong to this category and have been rendering their service for 15.8 years.

From the data given in Table 1, observed that all the CLS times are greater than or equal to 2.5. It is also observed that n=19, m=17, g=304 and k=236. The time at which left truncation of the data made is t_0 were t_0 is the cut off date from which the observations are taken and it is seen that $t_0 = l = p$. It may be noted that the times point happens to be observed number there at t_2 as suggested by McClean and Gribbin (1987) is as follows, Let there be n observations, where person i has x_i years service at t for i=1, 2, … , n.

Let

$$Y_i = \begin{cases} 1, & \text{if person } i \text{ is present at } t_2 \\ 0, & \text{if person } i \text{ has left by } t_2 \end{cases}$$

Then,

Prob (Y_i=1) = F ($x_i + t_2 - t_1$) / F(x_i) for i = 1, 2, , n
= p_i, Say

So, Y_i has a Binomial distributions with parameters 1 and p_i. The Y_i's are independent, So E[Y_i] = p_i and Var [Y_i] = p_i (1 – p_i). We assume that n is large, so by the central limit theorem the distribution of $Y = \sum_{i=1}^{n} Y_i$ is asymptotically normal, where

$$E(Y) = \sum_{i=1}^{n} p_i \text{ and } Var(Y) = \sum_{i=1}^{n} p_i (1 - p_i)$$

Then under the null hypothesis that the completed length of service distribution is F(.) we have that

$$\chi^2 = \frac{\left(Y - \sum_{i=1}^{n} p_i \right)^2}{\sum p_i (1 - p_i)}$$

has a Chi-square distribution with one degree of freedom. This is the test statistics for the goodness of fit test.

We have $t_2 - t_1$ = 0.56 years and

$$\Pr(Y_i = 1) = \left[\frac{x_i}{x_i + 0.56} \right]^{0.0432}$$

There are 17 persons in the case (ii) and 19 persons in the case (i) who left the organization by the time t_1. Hence, out of 566, there are 530 persons serving for the organization at time t_1. It is observed, by the time t_2, out of 530 persons, 16 have left the organization. Hence $Y = \sum_{i=1}^{530} Y_i = 514.$

The p value is 0.9733.

Conclusion

For the collected manpower data, a maximum likelihood estimate of Gamma distribution is obtained where left truncation and right censoring is observed. Numerical examples are provided for the particular case of Exponential distributions. In this paper, the estimation is done based on the data obtained for a long time span namely 15.8 years, say T and prediction for a shorter period namely 0.56 years, say t. Estimation and prediction is possible for different values of the T and t. The model developed in this paper provides a tool for assessing the incomplete manpower data both right censored and left truncated observations. The proposed methodology is applicable not only to the Tamilnadu Software Industry but also wider context in the other application areas.

ACKNOWLEDGEMENTS

The authors are indebted to Prof. R. Sathiyamoorthi for his help and encouragement in completing this paper. Thanks also due to U.G.C for their financial assistance under Rajiv Gandhi National Fellowship.

REFERENCES

Bartholomew, D. J. 1982. The Stochastic Model for Social Processes, 2nd ed., John Wiley and Sons, New York.

Bartholomew, D. J. and Forbes, A.F. 1979. Statistical Techniques for Manpower Planning, John Wiley and Sons. Chichester.

Kalbfleisch, J. D. and Prentice, R. C. 1980. The Statistical Analysis of Failure time data. John Wiley and Sons, New York.

Kaplan, E. L., and Meier, P.1958. Non-parametric Estimation form Incomplete Observations. *Journal of American Statistical Association,* 53:457-481.

Lane, K. F., and Andrew, J. E. 1955. A Method of Labour Turnover Analysis. J. *R. Statist. Soc.,* A118: 296-323.

Lowless, J. F. 1982. Statistical Models and Methods for Lifetime Data. John Wiley and Sons, New York.

McClean, S. I. 1976. A Comparison of the Log-Normal and Transition Models of Wastage, The Statistician, 25:281-294.

McClean, S. I. 1991. Manpower Planning Models and their Estimation: Theory and Methodology, *European Journal of Operational Research,* 51: 179-187.

McClean, S. I., and Gribbin, J. O. 1987. Estimation for Incomplete Manpower Data. *Appl. Stochastic Models and Data Analysis,* 3: 13-25.

Silcock, H. 1954. The Phenomenon of Labour Turnover. *J. R. Statist. Soc.,* A, 117: 429-440.

Susiganeshkumar, E., and Elangovan, R. 2009b. Analysis of Flows in a Manpower System with a Special Reference to Tamilnadu Software Industry. *Int. J. Agricult. Stat. Sci.,* 5: (2)629-640.

Susiganeshkumar, E., and Elangovan, R. 2009a. Predicting the Growth of Manpower for the Tamilnadu Software Industry (India). *Int. Journal of Physical Sciences*. 21(2): 377-392.

Turnbull, B. W. 1976. The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data. *J. R. Statistical Society,* 38: 290-295.

****