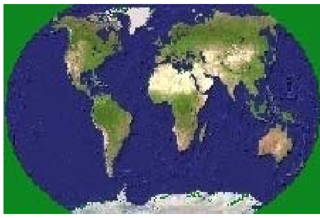
International Journal of Multidisciplinary Research and Development



International Journal of Multidisciplinary Research and Development

Volume: 2, Issue: 5, 364-367
May 2015
www.allsubjectjournal.com
e-ISSN: 2349-4182
p-ISSN: 2349-5979
Impact Factor: 3.762

J. Kalaivani
Department of Statistics,
Annamalai University,
Annamalai Nagar, India

R. Elangovan
Department of Statistics,
Annamalai University,
Annamalai Nagar, India

Correspondence:
J. Kalaivani
Department of Statistics,
Annamalai University,
Annamalai Nagar, India

A study on some estimation methods for incomplete manpower data

J. Kalaivani, R. Elangovan

Abstract
In manpower planning, one of the most important variables is completed length of service on leaving ajob, since it enables us to predict staff turnover. This variable is essentially the same as failure time, defined in the medical and reliability literature to be the time until a specified event occurs. In the medical literature this event is frequently death and in reliability theory, the time until a component breaks down. In manpower planning the data 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 munch work has been done on the parametric and non-parametric estimation of the survivor function using manpower data. In this paper, it is propose to discuss the various estimation methods for incomplete manpower data. Numerical illustration is also provided.

Keywords: Manpower Planning, Incomplete Manpower Data, Parametric and Non- Parametric Estimation, Survivor Function.

Introduction

Much attention has been paid to both non-parametric and parametric estimation for survival data with right censoring, particularly in the medical literature. The problem of estimating the life distribution based on incomplete manpower data is also an important aspect of study in the area of manpower planning. In manpower planning, one of the most important variables is duration until a specified event occurs. This is frequently the completed length of service until leaving a job, which enables us to predict staff turnover. However, planners are also interested in other duration such as length of service in a grade until promotion, or length of a spell of withdrawal from the labour force. It is also noted that, in manpower planning the completed length of service until leaving is of great interest, and here also the data are right censored since people are still in service when data collection ends. However, it often occurs that the data are also left truncated since people are already in service at the beginning of data collection. Lifetime data often come with a feature that creates special problems in the analysis of the incomplete manpower data. An observation is said to be right censored at L, if the exact value of the observation is not known but only that it is greater than or equal to L. Similarly, an observation is said to be left censored at L, if it is known only that the observation is less than or equal to L. Right censoring is very common in life time data, but left censoring is fairly rare. For such data much work has been done on both non-parametric and parametric estimation of the survivor function. It is worth mentioning at this juncture that the present thesis has taken up the paper by McClean and Gribbin (1987) McClean (1991) as a clue for developing many of the manpower planning models relating to estimation problems.

In Manpower planning, in an organization, wastage occurs mainly in two forms – voluntary or involuntary wastage. Voluntary wastage may be due to an individual's decision to quit the job or due to voluntary retirement whereas involuntary wastage may happen due to retrenchment, retirement or death of the person. The reasons for wastage provide additional information to Human Resource Development (HRD) about the leaving process of the personnel in the organization so that the organization, if need be, suitably modify the existing policies for the future. Wastage can be studied on the basis of an individual's exposure in his profession. Several authors including Bartholomew (1982), Bartholomew and Forbes (1979) have studied the concept of wastage and its relation to the Completed Length of Service (CLS).

The data on CLS are often incomplete due to left truncation as well as right censoring. For instance, 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 the study. The data were collected between March 2009, say t_0 (commencement of the period), and August 2014 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 Software Industry, refer to Susiganeshkumar and Elangovan (2009a), Susiganeshkumar and Elangovan (2009b).

The data on the CLS are of the following categories

i) Person who start his service on after t_0 and quits the organization on or before to the category called ' Complete data', denote as category 'A'.
ii) Person who was in service before t_0 itself and completes his service on or before t_1 belongs to the category 'left trucated' denote as category 'B'.
iii) Person who start his service for the organization on or after t_0 and continues in his job at t_1 belongs to the category 'right censored' denote as category 'C'.
iv) Person who was in service before t_0 itself and continues in the job at t_1 belongs to the category ' left truncated and right censored' denote as category 'D'.

Let as assuming that CLS follows Modified Weibull distribution. The maximum likelihood estimate of the parameter is obtained using the observations belonging to the four different categories defined above. The goodness of fit is also tested.

Parametric Estimation

The Modified Weibull distribution is a generalization of the exponential distribution that is appropriate for modeling lifetime having constant, strictly increasing, and strictly decreasing hazard functions. For a detailed study, refer to Sarhan, A.M, and Zain-din,M.,(2009).

Let T denotes the CLS follows the Modified Weibull distribution. The probability density function of T is given by

$$f(x;\alpha,\beta,\gamma)=(\alpha+\beta\gamma x^{\gamma-1})e^{-\alpha x-\beta x^{\gamma}}\ ,x>0 \qquad (1)$$

where $\beta>0$ and $\alpha>0$ are parameters sometimes referred to as the shape and scale parameters of the distribution.

The survivor and hazard functions are respectively,

$$s(t)=e^{-\alpha x-\beta x^{\gamma}} \qquad (2)$$

$$h(x;\alpha,\beta,\gamma,)=\alpha+\beta\gamma x^{\gamma-1},x>0 \qquad (3)$$

The hazard function (1) when $\gamma = 1$ is constant (2) $\gamma<1$ is decreasing, (3) $\gamma>1$the hazard function will be increasing.. For a detailed study refer to Sarhan, A.M, and Zain-din,M.,(2009).

Assume that there are n observatios $t_1,t_2,t_3,\ldots t_n$ which are complete (category A) and g observations right censored at $s_1,s_2,s_3,\ldots s_g$ (category C). There are m observed completed lengths of service (CLS) $r_1,r_2,r_3,\ldots r_m$ which are all left truncated at the same point (category B) and right censored observations $q_1,q_2,q_3\ldots\ldots q_k$ which are all at the same point p (category D). The parametric of the distribution is estimated by using the MLE method suggested by McClean and Gribbin (1987)

The likelihood function is given by

$$L=\left(\prod_{i=1}^{n} f(\ t_i)\right)\left(\prod_{i=1}^{g} F(s_{i)}\right)$$

subistute equation (2) and (3) we get

$$L=\prod_{i=1}^{n}(\alpha+\beta\gamma x_i^{\gamma-1})e^{-\alpha x_i-\beta x_i^{\gamma}}$$

The log likelihood function from censored observation becomes

$$\log L=\log\left\{\sum_{i=1}^{n}\left\{(\alpha+\beta\gamma x_i^{\gamma-1})e^{-\alpha x_i-\beta x_i^{\gamma}}\right\}\right\}$$

$$\log L=\left\{\sum_{i=1}^{n}\log\left\{(\alpha+\beta\gamma x_i^{\gamma-1})\right\}-\alpha\sum_{i=1}^{n}x_i-\beta\sum_{i=1}^{n}x_i^{\gamma}\right\}$$

Computing the first partial derivative of logL and setting the results equal zeros, we get the likelihood equations as in the following equations.

$$\sum_{i=1}^{n}\frac{1}{(\alpha+\beta\gamma x_i^{\gamma-1})}-\sum_{i=1}^{n}x_i=0, \qquad (4)$$

$$\sum_{i=1}^{n}\frac{\gamma x_i^{\gamma-1}}{(\alpha+\beta\gamma x_i^{\gamma-1})}-\sum_{i=1}^{n}x_i^{\gamma}=0, \qquad (5)$$

$$\sum_{i=1}^{n}\frac{x_i^{\gamma-1}\left(1+\gamma\log(\ x_i)\right)}{(\alpha+\beta\gamma x_i^{\gamma-1})}-\sum_{i=1}^{n}x_i^{\gamma}\log(\ x_i)=0. \qquad (6)$$

Using (4) and (5), we can derive β as a function of α and γ as in the form

$$\beta=g(\alpha,\gamma)=\frac{n-\alpha\sum_{i-1}^{n}x_i}{\sum_{i=1}^{n}x_i^{\gamma}}, \qquad (7)$$

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).

Differentiating gives the maximum likelihood,

$$\hat{\lambda}=\frac{n+m}{\sum_{i=1}^{n}\log t_i+\sum_{i=1}^{g}\log s_i+\sum_{i=1}^{m}\log r_i-\sum_{i=1}^{k}\log q_i-k\log p-(n+g)\log a} \qquad (8)$$

$$=\frac{Total\quad levers}{Total\quad observed\quad service\quad times}$$

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 Software Industry, the results were shown in the following Table 1.

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

Category A Complete data (t_i)	7.8, 5.5, 6.5, 9.6, 3.3, 4.7, 2.7, 10.11, 3.7, 5.2, 7.2, 9, 6.9, 6.5, 3.1, 4.2, 11.5, 12.4, 10.4 ,5.6,11.5
Category B Left truncated (r_i)	8.7, 10.2, 6.7, 5.5, 8.2, 4.6, 3.1, 4.5, 5.2, 5.1, 6.4, 5.3, 3.4, 2.7, 4.4, 7.2, 10.4,9.2
Category C Right censored (S_i)	13.4, 15.8, 15.8, 15.8, 15.8, 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, 12.1, 10.2, 13.2,13.2,13.2,13.2,12.1,10.2, 13.2, 13.2, 13.2, 13.2 ,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, 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, 11.3, 11.4, 11.3, 11.3, 11.3, 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, 15.8, 15.8, 15.8, 15.8, 15.8, 15.8, 15.8, 15.8, 15.8, 9.1, 7.6, 7.5, 7.5, 7.5, 7.5, 7.5, 7.1, 7.5, 7.6, 7.6, 7.6, 7.6, 7.6, 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, 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, 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, 11.2, 12.2, 12.2, 15.8, 15.8, 15.8, 15.8, 15.8, 15.8, 10.2, 10.3, 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, 6.3, 6.3, 6.3, 6.3, 6.3, 6.3, 4.6, 6, 4.6, 4.6, 1.2, 2.3
Category D Left truncated and Right censored (q_i)	There are 217 individuals who belong to this category and have been rendering their service for 14.7 years.

From the data given in Table 8.1, observed that all the CLS times are greater than or equal to 2.5. It is also observed that n=21, m=18, g=302 and k=217. 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 12.3 years since the commencement of the Tamilnadu Software Industry namely the origin. It is seen from the data set that

$$\sum_{i=1}^{m=18} \log r_i = 0.07, \quad \sum_{i=1}^{n=21} \log t_i = 1.98, \quad \sum_{i=1}^{k=217} \log q_i = 2.85 \quad and \quad \sum_{i=1}^{g=302} \log S_i = 2.54.$$

$\hat{\lambda}$ *is* estimated as 0.0217.

The Goodness of Fit Test

We want 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_o 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 observed number there at t_2. For each individual we may estimate his probability of surving to t_2 given that he is there at t_1 , and we develop a chi-squared test to take this into account as follows. Let there be *n* observations, where person *i* has *x;* years' service at *t* for *i* = 1, . . . , *n.*

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

Then

$$\text{Prob}\,(Y_i = 1) = F(x_i + t_2 - t_1)/S(x_i), \quad S(x_i)\, is\, the\, survivor\ function$$
$$= p_i\,, say$$

So Y_i has a binomial distribution with parameters 1 and p_i. The Y_i s are independent, so $E[Y_i] = p_i$ and $Var(Y_i) = p_i(l - p_i)$. 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

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

Then under the null hypothesis that the completed length of service distribution is Modified wiebull distribution. The test statistics $\chi^2 = \left(Y - \sum_{i=1}^{n} p_i\right)^2 \Big/ \sum p_i(1-p_i)$ has a chi-squared distribution with one degree of freedom. This test statistic for the goodness of fit test suggested by MeClean and Gribbin (1987).

We have $t_2 - t_1$ = 0.42 years and $\Pr(Y_i = 1) = \left[\dfrac{x_i}{x_i + 0.42}\right]^{0.0321}$

There are 21 persons in the case (ii) and 18 persons in the case (i) who left the organization by the time t_1. Hence, out of 519, there are 480 persons serving for the organization at time t_1. It is observed, by the time t_2, out of 480 persons, 15 have left the organization. Hence $Y = \sum_{i=1}^{480} Y_i = 465.$

The p value is 0.5421.

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 distribution. In this chapter, the estimation is done based on the data obtained for a long time span namely 14.7 years, say T and prediction for a shorter period namely 0.42 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 in wider context in the other application areas.

Reference

1. Bartholomew, D. J., A. F. Forbes, and S. I. McClean (1991). *Statistical Techniques for Manpower Planning. Wiley,* Chichester.
2. Bartholomew, D. J. (1982). *The Stochastic Model for Social Processes*. 3nd ed., John Wiley and Sons, New York.
3. Bartholomew, D. J., A. F. Forbes, and S. I. McClean (1991). *Statistical Techniques for Manpower Planning*. Wiley, Chichester.
4. Bartholomew, D. J., and A.F. Forbes (1979). *Statistical Techniques for Manpower Planning*. John Wiley and Sons, Chichester.
5. Biswas, S. (1988). *Stochastic Processes in Demography and Application.* Wiley Eastern Limited, Calcutta.
6. Biswas, S. (1996). *Statistical Techniques for Manpower Planning and Forecasting*. New age International Publisher, New Delhi.
7. Kaplan, E.L., and P. Meier. (1958). *Non-parametric Estimation from Incomplete Observations*. Journal of the American Statistical Association, Vol.53, pp.458-479.
8. McClean. S. I. and J. O. Gribbin. (1987). *Estimation for Incomplete Manpower Data.* Applied Stochastic Models and Data Analysis, Vol. 3, pp.337-341.
9. McClean. S. I. (1975). *A comparison of the lognormal and transition models of wastage*, the Statistician, Vol. 25, No. 4, pp. 281-293.
10. Susiganeshkumar, E., and R. Elangovan (2009a). *Predicting the growth of manpower for the Tamilnadu Software Industry (India)*. Int. Journal of Physical Sciences, Vol. 21(2) M, pp. 377-392.
11. Susiganeshkumar, E., and R. Elangovan (2009b). *Analysis of Flows in a Manpower System with a Special Reference to Tamilnadu Software Industry*. Int. J. Agricult. Stat. Sci., Vol. 5, No. 2, pp. 629-640.
12. Vijayasankar, N., Elangovan, R., and R. Sathiyamoorthy (2008). *A Stochastic Model for the Estimation of Expected time to Recruitment due to Attrition*. Int. J. Agricult. Stat. Sci., Vol. 4, No. 2, pp. 387-397.
13. Young, A., and G. Almond (1961). *Predictions Distributions of Staff.* Computer Journal, Vol. 3, pp.246-250.