Available Online at http://www.recentscientific.com

International Journal of Recent Scientific Research

International Journal of Recent Scientific Research
Vol. 4, Issue, 3, pp.185 -197, March, 2013

ISSN: 0976-3031

RESEARCH ARTICLE

PARAMETRIC AND NON-PARAMETRIC ESTIMATION OF INCOMPLETE MANPOWER DATA USING COX'S APPROACH

***Arulpavai, R and [2]Elangovan, R**

*Department of Statistics, Government Arts College, C.Mutlur, Chidambaram – 608 001
[12]Department of Statistics, Annamalai University, Annamalai Nagar – 608 002

ARTICLE INFO

Article History:

Received 12th January, 2012
Received in revised form 20th, January, 2012
Accepted 16th February, 2013
Published online 28th March, 2013

Key words:

Incomplete Manpower Data, Complete Length of Service, Kaplan and Meier's Estimate and Cox's Approach

ABSTRACT

In manpower planning the data are often incomplete due to left truncation as well as 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. Several models have been suggested to describe the internal and external movements of staff in a commercial or industrial organization relating to censored data. 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 most widely used distributions for completed length of service until leaving are the mixed exponential distribution and the lognormal distribution. The Weibull distribution is a particularly important life distribution and a large body of literature on statistical methods has evolved for it. In place of the Weibull distribution, it is often more convenient to work with the equivalent extreme value distribution when the observation are censored. In this paper a parametric and non-parametric estimate of the survivor function which extends Kaplan and Meier's estimate to include left truncation as well as right censoring has been discussed. A suitable model is also developed, to analyze and predict the pattern of manpower wastage and the basis of all possible individual characteristics responsible for the wastage on the basis of Cox's approach using Weibull and Extreme value distribution. A real industrial data has been used to validate the above models.

INTRODUCTION

Mathematicians and statistician have done much work on the development of models of manpower systems in the years since the paper by Seal (1945). Manpower planning had been commonly described as a process consisting of three elements (i) predicting the future demand for manpower; (ii) predicating the future supply of manpower; (iii) looking at policies to reconcile any difference between the results of (a) and (b), often known as "closing the manpower gap". Predicating demand may involve looking at productivity changes, technological changes, market forces and trends and the corporate strategy: predicating supply involves a knowledge of the current manpower stocks and looking at future recruitment, wastage, working conditions, promotion policies and labour market trends: closing the gap means examining training, remuneration, career planning, redundancies and further consideration of all the factors under the other headings. For a detailed study refer to Edwards (1983)

It is a common phenomenon that some personnel leave an organization after completing a certain period of services to that organization either voluntarily or due to death, retirement or termination, known as 'turnover' or 'wastage'. Wastage creates vacancies to be filled up either by promotion from the lower cadres or by direct recruitment or by a combination of both. The problem of manpower planners therefore, is to estimate the extent to which skilled or trained personnel are likely to leave over future points of time, given that the status quo viz., the present service condition and the personnel characteristics of the individuals is maintained, with a view to implement suitable measures for future recruitment, training . This kind of exercise is important in manpower planning, especially, in a situation where vacancies cannot be immediately filled up. Wastage of trained and experienced personnel cause loss to an organization in terms of money, time, efficiency and business and technological secrecy. Thus another major problem for manpower planners is to identify the individual characteristics of the personnel basis of all possible individual characteristics responsible for the wastage. This can possible throw some light into the trend in the recruitment process, especially, while leaving and recruitment may run simultaneously but vacancies may not be filled up immediately following an attrition or wastage.

* Corresponding author: **R. Arulpavai**
E-mail address: r_arulpavai@yahoo.co.in

In any 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). Some interesting results can also seen in Sathiyamoorthi et al. (2005), Elangovan et al. (2005), Anantharaj et al. (2006), Vijayasankar et al. (2008), Elangovan et al. (2008), Susiganeshkumar and Elangovan (2009a), Susiganeshkumar and Elangovan (2009b), Arivazhgan and Elangovan (2010), Susiganeshkumar and Elangovan (2012).

The propensity of leaving the job is quantified based on CLS only and it sometimes leads to unrealistic prediction of manpower attrition. In addition to CLS, at the micro level, other factors such as socio-economic indicators or personal and/or familial factors influence the propensity to leave the job. The Cox's regression model (1972) is being used to describe the manpower attrition by considering both CLS and personal cofactors where CLS is described by a quantitative variable while personal cofactors are described by qualitative variables. When both quantitative and qualitative characters are taken into account for studying manpower wastage and attrition the results are realistic and gives a meaningful prediction of manpower attrition.

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. It is often the case that failure time data is right censored, i.e. for some of the data we know only that failure takes place after a certain time but not exactly when it occurs. In the medical literature this corresponds to there being patients in the sample who are still alive when data collection is terminated, so we know only that their lifetime is greater than a certain value. For such data much work has been done on both non-parametric and parametric estimation of the survivor function. A maximum likelihood non-parametric estimate of the survivor function was derived by Kaplan and Meier (1958) and, as discussed by Kalbfleisch and Prentice (1980) this has been extended by various authors. Parametric models have also been fitted to right censored failure time data. In particular, maximum likelihood estimation for the exponential distribution is discussed by Kalbfleish and Prentice (1980) for medical data and by Tuma and Hannan (1979) for event history data in the sociological literature. Some interesting results can also be seen in McClean et. al. (1991).

For censored data, 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. A more general formula for obtaining a maximum likelihood estimate of the survivor functions has been found by Turnbull (1976) for arbitrarily grouped, censored and truncated data. Several models have been suggested to describe the internal and external movements of staff in a commercial or industrial organization. Of these, the Weibull model of Lane and Andrew (1955) which relates an employee's probability of leaving to his length of service, and the transition model of Young and Almond (1961), which considers the various grades in the company hierarchy, and the state of having left, to be the states of a Markov Chain, are among the simplest and easiest to apply. These models appear to give satisfactory results for most companies under normal conditions. In this paper a suitable model is developed, to analyze and predict the pattern of manpower wastage and the basis of all possible individual characteristics responsible for the wastage on the basis of incomplete manpower data and predicting the CLS using Cox's approach.

This paper is organized as follows. In Section 2 Cox's partial likelihood is discussed, the estimation of the longevity of service is discussed in Section 3. The parametric and non-parametric estimation which includes the Kaplan and Meier non-parametric estimation of the survival function and its confidence interval, the goodness of fit test has been discussed in Section 4 and Section 5. The factor affecting wastage is highlighted in Section 6. The result exhibiting goodness of fit test of prediction of leavers for different grades using Weibull and Extreme value distribution have been discussed in Section 7. A real data example is used to illustrate the Cox's approach is also discussed in Section 7. A summary of results is also highlighted in Section 8.

Cox's Partial Likelihood

The Cox's regression model (1975) based on the method of 'Partial likelihood' plays a very important role in analyzing the data in a more realistic way on survival or fertility on any other kind involving population characteristics using Stochastic models. The partial likelihood estimating the parameters by the method of maximum likelihood conjectured that the method would give estimates of β_1, $\beta_{2,\ \ldots,}$ β_p which would have otherwise the asymptotic properties of the maximum likelihood estimators. The instantaneous propensity of leaving a job (or profession) at time t is defined as the conditional probability of leaving a job (or profession) during an infinitesimal interval (t, t + dt) given that the person was in job till the time t. Denoting the hazard rate by h(t), we have

$$h(t)dt = \frac{f(t)dt}{R(t)} \qquad \ldots (2.1)$$

where R(t) is the Survival function or the probability of continuing the job at least upto a period t and f(t)dt is the probability of leaving the job between (t, t + dt). It can be shown that

$$R(t) = \exp\left[-\int_0^t h(\tau)d(\tau)\right] \text{ and } f(t) = \frac{d}{dt}\big(1\text{-}F(t)\big) = \frac{d}{dt}R(t) \qquad \ldots (2.2)$$

The Cox hazard model as indicated by (vide Gill (1984)).

$$h(t) = h_0(t)\exp\left(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_k X_k\right) \qquad \ldots (2.3)$$

Where $h_0(t)$ represents the hazard rate or the rate of propensity of leaving at time t purely on the consideration of time or CLS in the profession. $h_0(t)$ is called the base line hazard rate. The probability that the i-th person will leave the job at time t in (0, T) is given by

$$\frac{h_0(t)e^{\beta_1 X_{1i}+\beta_2 X_{2i}+\beta_3 X_{3i}}}{h_0(t)\sum_{i=1}^{n} e^{\beta_1 X_{1i}+\beta_2 X_{2i}+\beta_3 X_{3i}}}, \quad (i = 1,2,\cdots,n) \quad \text{... (2.4)}$$

where $X_{1i} + X_{2i} + X_{3i}$ are the scores of the covariates 1, 2 and 3 respectively of the i-th person. Note that the ratio in eqn. (2.4) is independent of t, the length of service. If we take the product of all such terms for all the professionals with serial number 1, 2, …, k, we get a simplified form of Cox's partial likelihood given by

$$P_L = \prod_{i=1}^{k} \frac{e^{\beta_1 X_{1i}+\beta_2 X_{2i}+\beta_3 X_{3i}}}{\sum_{i=1}^{k} e^{\beta_1 X_{1i}+\beta_2 X_{2i}+\beta_3 X_{3i}}} \quad \text{… (2.5)}$$

Maximizing P_L (or $\log P_L$) with respect to β_1, β_2, and β_3 respectively, we get three estimating equations for estimating β_1, β_2, and β_3 (assuming $e^{\beta_1 X_{1i}+\beta_2 X_{2i}+\beta_3 X_{3i}} \cong (1+\beta_1 X_{1i}+\beta_2 X_{2i}+\beta_3 X_{3i})$ approximately both in the numerator and denominator of eqution (6.5)) as follows:

$$\sum_{i=1}^{k} X_{1i} - k\,\frac{\sum_{i=1}^{k} X_{1i}(1+\beta_1 X_{1i}+\beta_2 X_{2i}+\beta_3 X_{3i})}{\sum_{i=1}^{k}(1+\beta_1 X_{1i}+\beta_2 X_{2i}+\beta_3 X_{3i})} = 0 \quad \text{… (2.6)}$$

$$\sum_{i=1}^{k} X_{2i} - k\,\frac{\sum_{i=1}^{k} X_{2i}(1+\beta_1 X_{1i}+\beta_2 X_{2i}+\beta_3 X_{3i})}{\sum_{i=1}^{k}(1+\beta_1 X_{1i}+\beta_2 X_{2i}+\beta_3 X_{3i})} = 0 \quad \text{… (2.7)}$$

$$\sum_{i=1}^{k} X_{3i} - k\,\frac{\sum_{i=1}^{k} X_{3i}(1+\beta_1 X_{1i}+\beta_2 X_{2i}+\beta_3 X_{3i})}{\sum_{i=1}^{k}(1+\beta_1 X_{1i}+\beta_2 X_{2i}+\beta_3 X_{3i})} = 0 \quad \text{… (2.8)}$$

Having estimated the parameters β_1, β_2, and β_3 relating to the covariates, the parameters affecting the leaving from profession for personal reason or covariates, we estimate the parameter of the CLS corresponding to the base line hazard function. For a detailed study on the subject in this direction, refer to Biswas (1988), Biswas and Adhikari (1992) and Fang and Li (2005).

Estimation of the Longevity of Service

Once the parameters viz., time dependent parameters concerning the CLS as well the parameters concerning the personal covariates β_1, β_2, and β_3 are estimated independently, the expected length of service can be obtained by the relationship between survival function R(t) and the expected or average duration of service given by L as

$$E(L) = \int_0^\infty R(t)dt = \int_0^\infty e^{\int_0^\infty -h(T)dT} \quad \text{… (3.1)}$$

$$= \int_0^\infty e^{-\int_o^t \exp(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3)h(T)dT} \quad \text{… (3.2)}$$

$$\left[\Theta\, R(t) = e^{-\int_0^t h(T)dT} \text{ and } h(t) = h_0(t)\exp(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3)\right]$$

Finally expected duration of service given by

$$E(L) = \int_0^\infty e^{\int_0^t -h(T)dT}$$

The proportion or the percentage of leavers between t and (t + 1) is accordingly given by

$$f(t) = \frac{d}{dt}(1 - R(t)) \text{ or } f(t) \times 100\% \quad \text{… (3.3)}$$

For a detailed study refer to Biswas (1996), Cox (1972), Cox (1975), Sathiyamoorthy *et al.* (2006d).

NON-PARAMETRIC ESTIMATION

We want to derive a non-parametric maximum likelihood estimate of the survivor function which uses all the data as fully as possible, including the incomplete left truncated and right censored data. The product limit estimate of the non-parametric survivor function was derived by Kaplan and Meier(1958) for right censored data. Using the notation of Kalbfleish and Prentice (1980), let $t_1 < t_2 < \ldots\ldots < t_k$ represent the observed failure times in a sample from a homogeneous population with survivor function F(t), where $t_0 = 0$. Suppose that d_j items fail at t_j, (j = 1, . . . , k) and m_j items are right censored in the interval $[t_j, t_{j+1})$. Let $n_j = (m_j + d_j) + \ldots + (m_k + d_k)$ be the number of items at risk at a time just prior to t_j. Then the maximum likelihood estimate of F(t) is given by

$$F(t) = \prod_{j|t_j<t} \left(\frac{n_j - d_j}{n_j}\right) \quad \text{… (4.1)}$$

So we make the conditional probability of failure at each t_j agree exactly with the observed conditional relative frequency of failure at t_j given by d_j / n_j. We now modify this result, to take into account the possibility of left truncation as well as right censoring. The data now include a_{jl} items which fail at t_j and are left truncated in $(t_{l-1}, t_l]$ at times s_{jlr}, where r = 1, . . . , a_{jl} and $l \leq j$. The data also include b_{jl} items censored in $[t_j, t_{j+1})$ at times c_{jlr}, which are left truncated in $[t_{l-1}, t_l)$ at times h_{jlpr} where r = 1, . . . , e_{jl} and p = 1, . . . , g_{jlr}, g_{jlr} is the number of items truncated in $(t_{l-1}, t_l]$ which are censored at c_{jlr} ; therefore

$$\sum_{r=1}^{e_{jt}} g_{jlr} = b_{jt}$$

This results in the likelihood function being modified to take left truncation into account by dividing by *F(x)* for each observation left truncated at *x*. Taking a similar approach to that of Kalbfleisch and Prentice this likelihood function is maximized by taking $F(s_{jlr}) = F(c_{jlr}) = F(h_{jlpr}) = F(t_j)$. Proceeding as for the Kaplan-Meier case, we now define

$$n_j = \left(d_j + \sum_{l=1}^{j} a_{jt} + m_j + \sum_{l=1}^{j} b_{jt}\right) + \ldots + \left(d_k + \sum_{l=1}^{j} a_{kl} + m_k + \sum_{l=1}^{j} b_{kl}\right)$$

as the number of items at risk just prior to t_j. Then

$$\hat{\lambda}_j = \frac{d_j + \sum_{l=1}^{j} a_{jl}}{n_j}, \; for\; j = 1, \ldots\ldots, k$$

(i.e. those who fail at t_j divided by those at risk immediately prior to t_j) is the maximum likelihood estimate of the hazard rate at t_j. The product limit estimate of the survivor function is therefore

$$\hat{F}(t) = \prod_{j|t_j<t} \frac{n_j - d_j - \sum_{l=1}^{j} a_{jl}}{n_j} \qquad \ldots (4.2)$$

which is maximum likelihood.

An asymptotic ninety five percent confidence interval for $\hat{F}(t)$, suitably transformed to ensure that value lie in [0, 1], is

$$\hat{F}(t)\exp(\pm 1.96\hat{S}(t)) \qquad \ldots (4.3)$$

where

$$\hat{S}(t)^2 = \sum_{l=1}^{j} \left(\frac{d_j + \sum_{l=1}^{j} a_{jl}}{n_j\left[n_j - \left(d_j + \sum_{l=1}^{j} a_{jl}\right)\right]} \right) \Bigg/ \sum_{j|t_j<t}^{j} \log\left(\frac{n_j - d_j + \sum_{l=1}^{j} a_{jl}}{n_j} \right)$$

This result is similar to that given by Kalbfleisch and Prentice (1980), Kaplan and Meier (1958). The above estimator may be obtained by first estimating the values missing as a result of left truncation and then using the, now complete, set of data to estimate the survivor function in the usual way. Forbes (1971) considered a similar problem when the data set was arbitrarily grouped. The above procedure is discussed in detail, the paper by McClean and Gribbin (1987). For a detailed study, refer to McClean *et. al.* (1991).

PARAMETRIC ESTIMATION

The Weibull Distribution

The 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 Leemisc (1995). The Weibull probability density function is given by

$$f(t) = \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1} \exp\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right] \quad t \geq 0 \qquad \ldots (5.1)$$

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,

$$s(t) = \exp\left[-(\alpha t)^{\beta}\right] \qquad \ldots (5.2)$$

$$h(t) = \beta(\alpha t)^{\beta-1} \qquad \ldots (5.3)$$

The hazard function approaches zero from infinity for $\beta < 1$, is constant for $\beta = 1$ the exponential case, and increase from zero when $\beta > 1$. For a detailed study refer to Leemisc (1995) and Lawless (1982).

The Extreme Value Distribution

In place of the Weibull distribution, it is often more convenient to work with the equivalent extreme value distribution with p.d.f.

$$f(x; u, b) = \frac{1}{b}\exp\left(\frac{x-u}{b}\right)\exp\left(-\exp\left(\frac{x-u}{b}\right)\right) \qquad \ldots (5.4)$$

where u $(-\infty < u < \infty)$ and $b (b > 0)$ are parameters. If T has probability density function given in equ. (5.1), then X=log T has p.d.f. given in equ. (5.4), with $u = \log \alpha$ and $b = \beta^{-1}$. The main convenience in working with the extreme value distribution stems from the fact that u and b are location and scale parameters. Any results derived in terms of one distribution are easily transferred to the other. The Weibull distribution is a particularly important life distribution and a large body of literature on statistical methods evolved for it. The survivor and hazard functions are respectively,

$$s(t) = 1 - \exp\left(-\exp\left(\frac{x-u}{b}\right)\right) \qquad \ldots (5.5)$$

$$h(t) = \frac{1}{b}\exp\left(\frac{x-u}{b}\right) \qquad \dots (5.6)$$

For a detailed study refer to Kalbfleisch (1980).

Maximum Likelihood Estimation

Let T_i represents the lifetime and L_i the fixed censoring time of the *i* th individual in a random sample of *n* individuals. One observes only $t_i = min\ (T_i, L_i)$ and whether the observation is a lifetime or a censoring time. The T_i 's are assumed to have a Weibull distribution or, equivalently, $X_i = log\ T_i$ has an extreme value distribution, with parameter *u* and *b*. Let $\eta_i = \log L_i,\ x_i = \log t_i$ and $\delta_i = 1$ or 0, according to whether $t_i = T_i$ or $t_i = L_i$, respectively;

The likelihood function is given by

$$L = \prod_{i=1}^{n} f(t_i)^{\delta_i} S(L_i)^{1-\delta_i} \qquad \dots (5.7)$$

where t_i is a lifetime, L_i is a censoring time and

$$\delta_i = \begin{cases} 1 \text{ is indicidual s i's lifetime is observed} \\ 0 \text{ if it is censored} \end{cases}$$

By using equ. (5.7), the likelihood function becomes

$$L(u,b) = \prod_{i=1}^{n}\left[\frac{1}{b}\exp\left(\frac{x_i-u}{b} - e^{(x_i-u)/b}\right)\right]^{\delta_i}\left[\exp\left(-e^{(\eta_i-u)/b}\right)\right]^{1-\delta_i} \qquad \dots (5.8)$$

Let $r = \sum \delta_i$ denote the number of observed.

We have

$$\log L(u,b) = -r\log b + \sum_{i\in D}\frac{x_i-u}{b} - \sum_{i=1}^{n}\exp\left(\frac{x_i-u}{b}\right) \qquad \dots (5.9)$$

and thus

$$\frac{\partial \log L(u,b)}{\partial u} = -\frac{r}{b} + \frac{1}{b}\sum_{i=1}^{n}\exp\left(\frac{x_i-u}{b}\right) \qquad \dots (5.10)$$

$$\frac{\partial \log L(u,b)}{\partial b} = -\frac{r}{b} - \frac{1}{b}\sum_{i=1}^{r}\frac{x_i-u}{b} + \frac{1}{b}\sum_{i=1}^{r}\frac{x_i-u}{b}\exp\left(\frac{x_i-u}{b}\right) \qquad \dots (5.11)$$

The m.l.e's $\hat{u}$ and $\hat{b}$ can be obtained by simultaneously solving

$$\frac{\partial \log L(u,b)}{\partial u} = 0 \text{ and } \frac{\partial \log L(u,b)}{\partial b} = 0$$

One convenient way do to this is to note that setting (5.10) equal to 0 gives

$$\exp(\hat{u}) = \frac{1}{r}\sum_{i=1}^{n}\exp\left(\frac{x_i}{\hat{b}}\right)^{\hat{b}} \qquad \dots (5.12)$$

Substituting this into $\frac{\partial \log L(u,b)}{\partial b} = 0$, we get

$$\sum_{i=1}^{r} x_i \exp\left(\frac{x_i}{\hat{b}}\right) \Big/ \sum_{i=1}^{r}\exp\left(\frac{x_i}{\hat{b}}\right) - \hat{b} - \left(\frac{1}{r}\right)\sum_{i=1}^{r} x_i = 0 \qquad \dots (5.13)$$

To find the $\hat{u}$ and $\hat{b}$ one can therefore determine $\hat{b}$ as the solution to (5.13) and then obtain $\hat{u}$ from (5.12). Since (5.13) cannot solve analytically for $\hat{b}$. The equation (5.13) can be solved iteratively using Newton's method for $\hat{b}$, then $\hat{u}$ calculated from eqn. (5.12).

The m.l.e.'s of the Weibull parameters β and α are $\hat{\alpha}$ = exp $\hat{u}$ and $\hat{\beta} = \hat{b}^{-1}$. If desired, the maximum likelihood equations (5.12) and (5.13) can be written in Weibull form and solved directly for $\hat{\beta}$ and $\hat{\alpha}$. If desired, the maximum likelihood equation (5.12) and (5.13) can be written in Weibull and solved directly for $\hat{\alpha}$ and $\hat{\beta}$. The log likelihood and maximum likelihood equations are respectively.

$$\log L(\alpha,\beta) = r\log\beta - r\beta\log\alpha + (\beta-1)\sum_{i\in D}\log t_i - \sum_{i=1}^{n}\left(\frac{t_i}{\alpha}\right)^{\beta} \qquad \dots (5.14)$$

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*.

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

Then

$$\text{Prob}(Y_i = 1) = F(x_i + t_2 - t_1)/F(x_i),\ for\ i = 1,2,\dots.n$$
$$= p_i$$

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 *F(.)* we have that

$$\left(Y - \sum_{i=1}^{n} p_i\right)^2 \Big/ \sum p_i(1-p_i) \qquad \dots (5.15)$$

has a chi-squared distribution with one degree of freedom. This is the test statistic for the goodness of fit test suggested by MeClean and Gribbin (1987). One year Prediction and Estimation period from different grades and its goodness of fit, significance level are shown in Table 7.4, Table 7.5 and Table 7.6. The median and expected length of service for the different grades in the case of Weibull distribution, Extreme value distribution and Kaplan-Meier are shown on Table 7.7.

TESTING THE MODEL

As an Estimator

Let N_{ij} be the number of people starting in year Y_i who are still in service in year Y_j (i = 1, …,n; j = i,…,n). Then N_{ii} people are recruited in year Y_i and N_{in} of these are still in service in year Y_n. Let p_i be the probability of surviving *i* years in the firm, and $\hat{p}_i$ an estimate of p_i obtained using either the Weibull or Extreme value models. Then, since $N_{in} \sim$ Binomial (N_{ii} , p_{n-i}), we have

$$\left(N_{in} - N_{ii}\hat{p}_{n-i}\right)^2 / \left\{N_{ii}\hat{p}_{n-i}\left(1-\hat{p}_{n-i}\right)\right\} \sim \chi^2(1)$$

Then, since the behaviour of each year's is independent of other, we have

$$\sum_{i=1}^{n-1}\left(N_{in}-N_{ii}\hat{p}_{n-i}\right)^2 / \left\{N_{ii}\hat{p}_{n-i}\left(1-\hat{p}_{n-i}\right)\right\} \sim \chi^2(n-1) \quad \text{… (5.16)}$$

This statistic can therefore be used to test the set $\{\hat{p}_j\}$ as an estimator for $\{p_j\}$ provided that the leaving probabilities are independent of time. The test statistic is suggested by McClean (1975).

As a Predictor

The model can also be tested as a predictor if the number of survivors from the last year, Y_n, used in estimation, to the following year, Y_{n+1}, is known. Thus if N_{in} is the number of people recruited in year Y_i who are in service in year Y_n and M_i is the observed number of these who survive one more year to Y_{n+1}, them M_i has a binomial distribution $B(N_{in}, r_i)$ where $r_i = p_{n-i+1} / P_{n-i}$. Therefore we have

$$\left(M_i - N_{in} r_i\right)^2 / \left\{N_{in} r_i \left(1-r_i\right)\right\} \sim \chi^2(1)$$

and since the N_{in} are survivors from different entry and therefore independent it follows that

$$\sum_{i=1}^{k}\left(M_i - N_{in} r_i\right)^2 / \left\{N_{in} r_i \left(1-r_i\right)\right\} \sim \chi^2(k) \quad \text{… (5.17)}$$

and this statistic can be used as a measure of the accuracy of prediction of the model. Obviously this is not just restricted to predictions for one year ahead and a similar statistic can be determined for testing predictions for any number of years ahead. The test statistic is suggested by McClean (1975).

Factors Affecting Wastage

Hedberg (1961) suggested that wastage is dependent on age, tenure, skill and responsibility, and sex. It is also reasonable that, within a company, or different functions, may vary. Each company was therefore tested to see if its wastage was dependent on each of tenure, sex, qualification and location, where appropriate this was done by means of an χ^2 test for non-association of 2xk contingency table formed by the total stayers and leavers, over a period of six years, and the classes of each factor considered as suggested by McClean et al. (1991).

Classes	1 2 3 4 . . . k	Total
Stayers	s_1 s_2 s_3 s_4 s_k	$S = \Sigma\, s_k$
Leavers	l_1 l_2 l_3 l_4 l_k	$L = \Sigma\, l_k$
		$N = S+L$

For non-association, the probability of leaving is independent of class, the test statistic is given by

$$\sum_{i=1}^{k}\left[\frac{s_i - S(s_i+l_i)}{N}\right]^2 \left[\frac{S}{N}(s_i+l_i)\left(1-\frac{S}{N}\right)\right] \sim \chi^2(k-1)\,d.f. \quad \text{… (6.1)}$$

The test statistics is suggested by McClean (1975).

RESULTS

Goodness of fit of prediction of leavers for different grades

The goodness of fit test of prediction of leavers for the grades Team leader, SW Engineer and System Administrator for the Estimation period and Prediction period using test statistics discussed in eqn. (5.15) of section 5.4 are given in Table 7.1, Table 7.2 and Table 7.3.

The estimator of the survivor function and confidence interval for different years using Kaplan Meier product limit estimator, Weibull distribution and Extreme value distribution as discussed in eqn. (4.2), eqn. (4.3), eqn. (5.1) and eqn. (5.2) of section 4 and section 5 respectively, and are shown in Fig. 7.1, Fig. 7.2 and Fig. 7.3 for different grades.

Table 7.1 Goodness of fit of prediction of leavers for Team leader

Estimation Period	Prediction Period	Kaplan Meier	Weibull Distribution	Extreme Value Distribution
100	100	0.4	1.6	14.1**
100	290	0.2	41.3**	64.6**
100	650	5.2*	210.0**	118.4**
290	100	0.2	22.2**	37.0**
290	290	0.1	0.3	1.4
290	650	3.0	103.2**	27.4**
650	100	0.0	0.2	2.1
650	290	0.2	32.5**	7.6**
650	650	12.2**	309.2**	89.9**

* significant at 5% level

** significant at 1% level

CLS as Weibull and Extreme value distribution

The χ^2 values for testing estimation as in eqn. (5.7) for Weibull and Extreme value distribution given in table 7.8 for the selected to companies of Tamilnadu. The χ^2 values for testing the prediction as in eqn. (5.8) for the Weibull and Extreme value distribution are given in table 7.9 for the selected 10 companies of Tamilnadu Cognizant (CTS), Tata consultancy service (TCS), Wipro, Polaris, HCL, Infosys, IC Info Tech, Larsen & Toubro Infotech, Ramco Systems Ltd, Sutherland Global Services Pvt. Ltd.

Table 7.2 Goodness of fit of prediction of leavers for Software Engineer

Estimation Period	Prediction Period	Kaplan Meier	Weibull Distribution	Extreme Value Distribution
100	100	1.9	3.2	0.2
100	290	2.2	3.1	5.5*
100	650	0.6	2.0	38.5**
290	100	0.4	0.1	0.2
290	290	20.1**	1.3	13.0**
290	650	47.5**	0.0	16.1**
650	100	0.4	4.4*	1.4
650	290	0.5	0.3	5.1*
650	650	1.9	3.9*	21.7**

* significant at 5% level

** significant at 1% level

Table 7.3 Goodness of fit of prediction of leavers for System Administrator

Estimation Period	Predictio n Period	Kaplan Meier	Weibull Distribution	Extreme Value Distribution
100	100	0.9	0.3	3.9*
100	290	0.5	3.4	8.4**
100	650	0.2	25.8**	13.0**
290	100	0.3	0.1	1.5
290	290	1.2	2.5	10.9**
290	650	0.2	6.9**	5.2*
650	100	0.0	0.3	0.9
650	290	1.3	1.8	5.2*
650	650	2.4	3.7	9.5**

* significant at 5% level

** significant at 1% level

Factors Affecting Wastage

The χ^2 test for non-association of 2xk contingency table formed by the total stayers and leavers over a period of six years, and the classes of each factor given in table 7.10, for the selected 10 companies of Tamilnadu software namely Cognizant (CTS), Tata consultancy service (TCS), Wipro, Polaris, HCL, Infosys, IC Info Tech, Larsen & Toubro Infotech, Ramco Systems Ltd, Sutherland Global Services Pvt. Ltd. Stayers and Leavers over a period of six years 2006-2011 given in table 7.10.

$$\chi^2=\sum_{i=1}^{k}\left[\frac{s_i-S(s_i+l_i)}{N}\right]^2\left[\frac{S}{N}(s_i+l_i)\left(1-\frac{S}{N}\right)\right]=298.74^{**}$$

** significant at 1% level.

COX'S APPROACH

Case (i): CLS as Weibull Distribution

The propenstative to leave the job based on four personal covariate collected from the data is given in table 7.11.

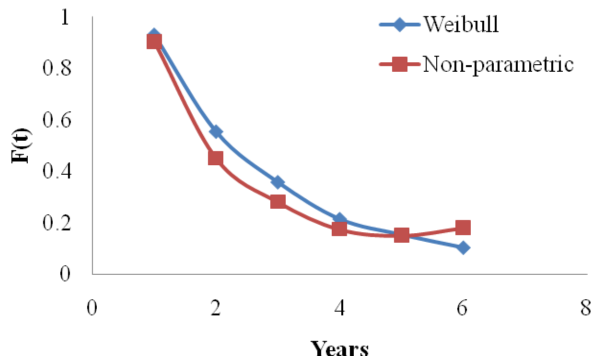
$$X_1=\begin{cases}1,\text{The job or workplace was not as expected}\\0,\text{otherwise}\end{cases}$$

$$X_2=\begin{cases}1,\text{Feeling devalued and unrecognizeds}\\0,\text{otherwise}\end{cases}$$

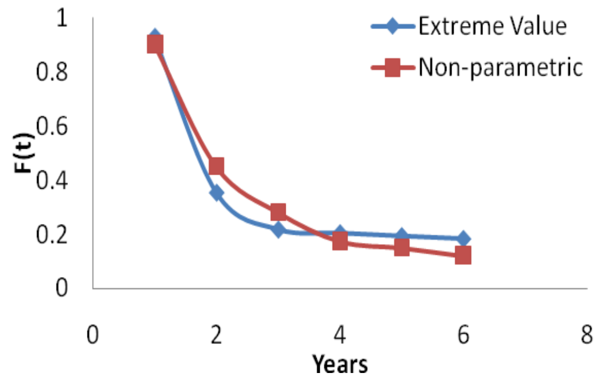
$$X_3=\begin{cases}1,\text{Stress from overwork and work - life imbalance}\\0,\text{otherwise}\end{cases}$$

$$X_4=\begin{cases}1,\text{Shifting Whims/Strategic Priorities}\\0,\text{otherwise}\end{cases}$$

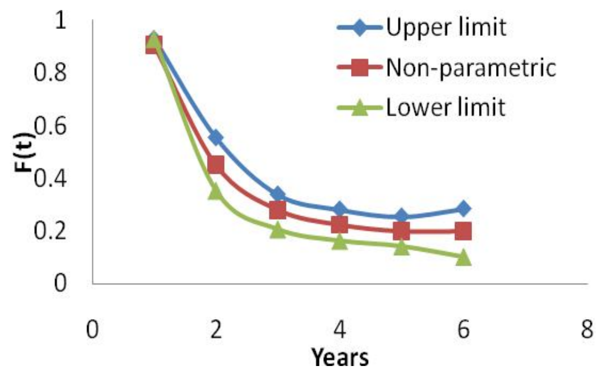
The hazard rates for the Weibull distribution are given in the Table 7.12.



(a) Weibull distribution and non-parametric



(b) Extreme value distribution and non-parametric

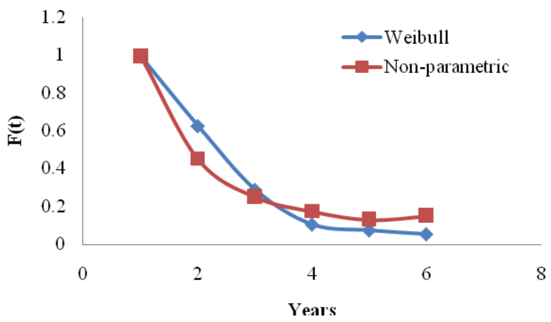


(c) Non-parametric survivor function: Upper limit and Lower limit

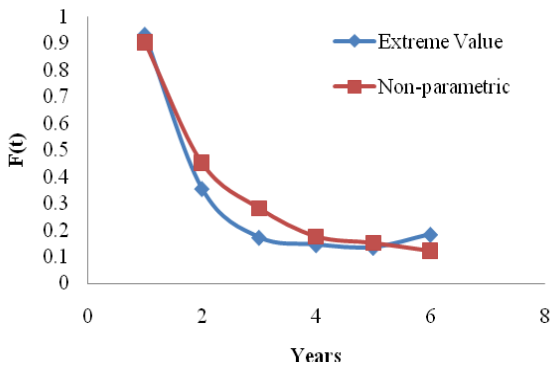
Fig. 7.1 Estimates of the survivor functions for Weibull distribution; non-parametric, Extreme value distribution; non-parametric, 95 % confidence interval for Kaplan-Meier estimate, its upper and lower confidence limits for Grade 1 (Team Leader).

Using the above data set, the parameter of the Weibull distribution given in eqn. (5.1), using the Maximum likelihood estimates in Section 5.3 and are estimated by using the Newton-Raphson method with the Statistical Analysis and

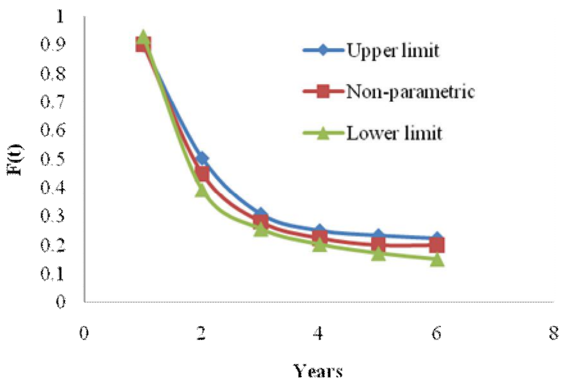
System (SAS) package, $\hat{\alpha} = 1.4123$, $\hat{\beta} = 0.5147$. After the estimating the parameters of the Weibull distribution the The hazard rates for the Weibull distribution are given in table 7.12.



(a) Weibull distribution and non-parametric



(b) Extreme value distribution and non-parametric



(c) Non-parametric survivor function: Upper limit and Lower limit

Fig. 7.2 Estimates of the survivor functions for Weibull distribution; non-parametric, Extreme value distribution; non-parametric, 95 % confidence interval for Kaplan-Meier estimate, its upper and lower confidence limits for Grade 2 (Software Engineer).

Table 7.4 One year prediction and estimation periods for Grade 1 (Team leader)

	Observed	Expected Kaplan-Mirer	Expected Weibull	Excepted Extreme Value
Leavers	359	362.537	558.498	379.428
Stayers	4005	4020.453	3806.825	3986.022
χ^2 with 1 d. f.		0.09903	81.57814**	1.190179

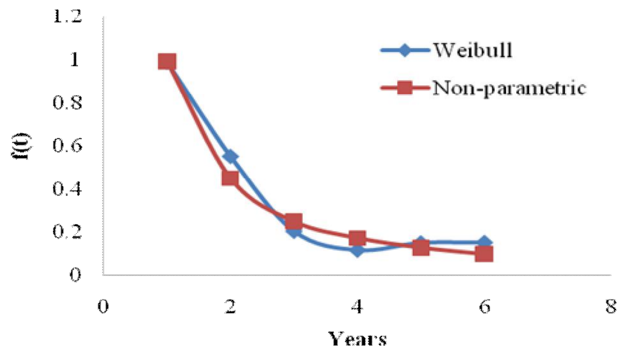
For Weibull distribution, $\alpha =$ 1.4123, β = 0.5147

For Extreme value distribution, b=3.5742, u=2.7439

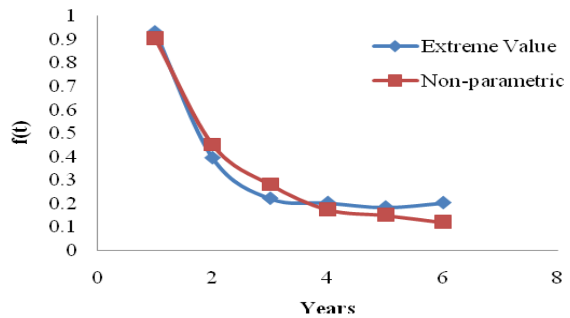
** - 1% level significant

Fig. 7.3 Estimates of the survivor functions for Weibull distribution; non-parametric, Extreme value distribution; non-parametric, 95 % confidence interval for Kaplan-Meier estimate, its upper and lower confidence limits for Grade 3 (System Administrator).

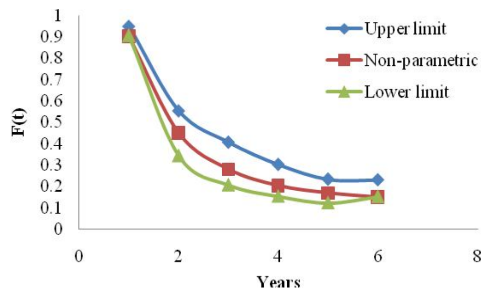
parameters $\beta_i = 1,2,3$ given in eqn. (2.6), eqn. (2.7) and eqn. (2.8) relating to the covariates are estimated as $\hat{\beta}_i = 1,2,3$ (0.6348, 0.1289, 3.1846).



(a) Weibull distribution and non-parametric



(b) Extreme value distribution and non-parametric



(c) Non-parametric survivor function: Upper limit and Lower limit

Fig. 7.3 Estimates of the survivor functions for Weibull distribution; non-parametric, Extreme value distribution; non-parametric, 95 % confidence interval for Kaplan-Meier estimate, its upper and lower confidence limits for Grade 3 (System Administrator).

Table 7.5 One year prediction and estimation periods for Grade 2 (Software Engineer)

	Observed	Expected Kaplan-Mirer	Expected Weibull	Excepted Extreme Value
Leavers	1290	1386.228	1709.1	1376.724
Stayers	6484	6388.652	6065.78	6398.156
χ^2 with 1 d. f.		8.102903**	131.6056**	6.614775*

For Weibull distribution, $\alpha =$ 1, β = 0.3

For Extreme value distribution, b=3, u=1.5

* - 5% level significant

** - 1% level significant

The probability of leaving the job upto 6 years of service, given the personal covariates (0,0,0,0), (0,0,0,1), (0,0,1,0), (0,0,1,1), (0,1,0,0), (0,1,0,1), (0,1,1,0), (0,1,1,1), (1,0,0,0), (1,0,0,1), (1,0,1,0), (1,0,1,1), (1,1,0,0), (1,1,0,1), (1,1,1,0), (1,1,1,1) are given in the following table 7.13, table 7.14 as per the above estimates.

Case (ii): CLS as Extreme Value distribution

For the CLS distributed as Extreme value distribution given in eqn. (5.4), and the Maximum Likelihood estimates given in section 5.3 and are estimated by using the Newton-Raphson method with the Statistical Analysis and System (SAS)

Table 7.6 One year prediction and estimation periods for Grade 3 (System Administrator)

	Observed	Expected Kaplan-Mirer	Expected Weibull	Excepted Extreme Value
Leavers	239	265.416	316.368	278.352
Stayers	2926	2899.235	2847.26	2886.14
χ^2 with 1 d. f.		2.876187	21.09792**	6.113887*

For Weibull distribution, $\alpha = 0.5$, $\beta = 1.5$

For Extreme value distribution, b=5.5, u=2.5

** - 1% level significant

Table 7.7 Median and expected lengths of service for one year estimation and prediction

Grade	Median Kaplan-Mirer	Median Weibull	Median Extreme Value	E(T) Kaplan -Mirer	E(T) Weibull	E(T) Extreme Value
1	0.9564	0.3971	4.0539	1.3425	0.4685	4.8069
2	1.4534	0.2079	2.5995	1.9215	0.3000	3.2316
3	2.8745	0.7206	4.5158	3.1247	3.0000	5.6746

Table 7.8 χ^2 values for testing estimation as in equ. (5.7)

Company	Weibull		Extreme Value	
	χ^2	d.f.	χ^2	d.f.
1	19.4362**	7	22.9272**	9
2	16.1313*	6	16.1623**	5
3	14.4374*	5	27.6143**	6
4	10.7315*	3	27.5484**	5
5	8.2623	4	18.9055**	4
6	6.8314	3	16.0286**	3
7	5.7175	2	14.7357**	2
8	3.7912	4	14.0382**	3
9	1.0532	3	12.8391*	4
10	3.7312	2	5.4304	3

* significant at 5% level

** significant at 1% level

Table 7.9 χ^2 values for testing prediction as in equ. (5.8)

Company	Weibull		Extreme Value	
	χ^2	d.f.	χ^2	d.f.
1	10.1412	9	15.1411	8
2	7.3522	5	12.3522	6
3	5.4533	7	15.5513	9
4	4.4321	13	14.5314*	7
5	3.4542	10	13.4582	7
6	17.4451	11	19.4563*	10
7	19.4622**	5	21.4724**	7
8	12.4733	7	15.4845	8
9	15.4744	12	19.4326*	10
10	6.4521	8	10.8474	6

* significant at 5% level

** significant at 1% level

Table 7.10 Stayers and Leavers over a period of six years 2006 – 2011

Classes	1	2	3	4	5	6	Total
Stayers	1158	1437	1528	1728	1324	1829	9004
Leavers	230	241	304	344	264	324	1707
							10711

Table 7.11 Data set with three Personal Covariates

Period in year	No. of persons left	Covariate
2006-07	25	(1,1,1,1) (1,0,1,1) (1,0,0,1) (1,0,0,0) (1,1,1,0) (1,1,0,0) (1,1,1,1) (1,0,0,0) (0,0,1,1) (0,0,0,1) (0,0,0,0) (0,1,1,1) (0,1,1,0) (0,0,0,0) (0,0,0,1) (0,0,0,0) (0,1,0,0) (0,0,1,0) (0,0,0,0) (0,0,0,0) (0,0,1,1) (0,1,1,0) (0,1,1,1) (0,0,0,0) (0,0,0,0)
2007-08	36	(0,0,0,0) (0,0,1,0) (0,0,0,0) (1,0,0,0) (0,1,0,0) (0,0,0,0) (0,0,0,0) (0,0,0,0) (0,1,1,0) (1,1,1,0) (0,0,0,0) (0,0,0,0) (0,1,0,0) (0,0,0,0) (0,0,0,0) (0,1,1,0) (1,1,1,0) (0,0,0,0) (0,0,0,0) (0,0,0,0) (0,1,0,0) (1,1,0,0) (0,1,1,0) (1,0,1,0) (0,1,0,0) (0,0,0,0) (0,0,0,0) (0,0,0,0) (0,0,1,0) (0,1,1,0) (1,1,1,0) (1,0,0,0) (1,1,0,0) (0,0,0,0) (0,0,0,0) (0,0,1,0)
2008-09	79	(1,1,1,1) (1,0,0,0) (1,1,0,1) (1,0,1,1) (1,1,1,0) (1,0,0,1) (1,0,0,0) (1,0,1,0) (1,1,0,0) (0,0,0,0) (1,1,0,1) (1,1,1,0) (1,0,0,0) (1,0,0,1) (1,0,0,0) (1,1,0,0) (1,0,1,0) (1,0,0,0) (0,0,0,0) (1,0,0,0) (1,0,1,1) (1,1,1,1) (0,0,0,0) (0,0,0,0) (1,0,1,0) (1,0,0,0) (1,1,1,0) (1,0,1,1) (1,1,0,1) (1,0,1,0) (1,1,1,1) (1,0,0,0) (0,0,0,0) (0,0,0,0) (0,0,0,1) (0,0,1,1) (1,1,1,1) (1,1,0,0) (1,1,1,0) (1,0,0,0) (0,0,0,0) (0,0,0,0) (0,0,0,1) (1,0,1,1) (1,0,1,0) (1,0,0,0) (1,0,0,0) (1,0,0,1) (1,0,0,0) (1,1,0,0) (1,0,1,0) (1,0,0,0) (0,0,0,0) (1,0,1,1) (1,1,1,0) (1,0,0,0) (1,0,0,0) (1,0,1,0) (1,1,0,1) (1,0,1,0) (1,0,0,1) (1,1,1,1) (1,0,0,0) (1,0,1,1) (1,0,0,1) (1,1,0,1) (1,1,1,0) (1,0,0,0) (1,0,1,0) (1,1,0,0) (1,0,0,0) (1,1,0,1) (1,1,1,0) (1,1,1,1) (1,0,0,0) (1,1,0,1) (1,0,1,1) (1,1,1,0) (1,0,0,1)
2009-10	43	(0,0,0,0) (0,0,0,1) (0,0,0,0) (0,1,0,0) (0,0,1,0) (0,0,0,0) (0,0,0,0) (0,0,1,1) (0,1,1,0) (0,0,0,0) (0,0,0,0) (0,0,1,0) (0,1,0,1) (0,0,1,0) (0,0,0,1) (0,0,1,1) (0,1,0,0) (0,1,1,0) (0,0,0,0) (0,0,1,0) (0,1,1,1) (0,0,0,0) (0,1,0,1) (0,0,1,1) (0,1,1,0) (0,0,0,1) (0,0,0,0) (0,0,1,0) (0,1,0,0) (0,1,0,1) (0,0,0,0) (0,0,0,1) (0,1,0,0) (0,0,1,0) (0,0,0,0) (0,0,1,1) (0,1,1,0) (0,0,0,0) (0,0,0,0) (0,0,1,0) (0,1,0,1) (0,0,1,0) (0,0,0,1)
2010-11	18	(0,0,1,1) (0,1,1,0) (0,0,0,1) (0,0,0,0) (0,0,1,0) (0,1,0,0) (0,1,0,1) (0,0,0,0) (0,0,0,1) (0,1,0,0) (0,0,1,0) (0,0,0,0) (0,0,1,1) (0,1,1,0) (0,0,0,0) (0,0,1,0) (0,1,0,1) (0,0,1,0)
64 persons who did not have up to 6 years of service		(0,0,0,0) (0,0,0,1) (0,0,0,0) (0,1,0,0) (0,0,1,0) (0,0,0,0) (0,0,0,0) (0,0,1,1) (0,1,1,0) (0,0,0,0) (0,0,0,0) (0,0,1,0) (0,1,0,1) (0,0,1,0) (0,0,0,1) (0,0,1,1) (0,1,0,0) (0,1,1,0) (0,0,0,0) (0,0,1,0) (0,1,0,1) (0,0,1,0) (0,0,0,1) (0,0,1,1) (0,1,0,0) (0,1,1,0) (0,0,0,0) (0,1,1,1) (0,0,0,0) (0,1,0,1) (0,0,1,1) (0,1,1,0) (0,0,0,1) (0,0,0,0) (0,0,1,0) (0,1,0,0) (0,1,0,1) (0,0,0,0) (0,0,0,1) (0,1,0,0) (0,0,1,0) (0,0,0,0) (0,0,1,1) (0,1,1,0) (0,0,0,0) (0,0,0,0) (0,0,1,0) (0,1,0,1) (0,0,1,0) (0,0,0,1) (0,0,1,1) (0,1,0,0) (0,1,1,0) (1,0,0,0) (0,0,1,0) (0,1,0,0) (0,0,0,0) (0,1,0,0) (0,1,1,0) (0,0,0,0) (0,1,1,0) (0,0,0,1) (0,0,0,0) (0,0,1,0)

Table 7.12 Hazard rate for the Weibull distribution

t	1	2	3	4	5	6
$h_0(t)$	0.4353	0.3109	0.2554	0.2221	0.1993	0.1824

Table 7.15 Hazard rate for the Extreme Value distribution

T	1	2	3	4	5	6
$h_0(t)$	0.1718	0.2272	0.3006	0.3976	0.5260	0.6958

Table 7.13 Probability of Leaving a Job within 6 Years of Service

Personal Covariate	**Up to 1 Yr**	**Between 1 -2 Yrs**	**Between 2 – 3 Yrs**	**Between 3 – 4 Yrs**	**Between 4 – 5 Yrs**	**Between 5 – 6 Yrs**
(1,1,1,1)	1.0000	1.0000	0.8177	1.0000	1.0000	1.0000
(1,1,1,0)	0.8177	0.8177	1.0000	1.0000	1.0000	0.8177
(1,1,0,1)	1.0000	0.2829	0.3141	0.4072	0.5635	0.6822
(1,1,0,0)	0.7418	1.0000	1.0000	1.0000	1.0000	0.8177
(1,0,1,1)	0.7613	0.8166	1.0000	0.8177	1.0000	1.0000
(1,0,1,0)	0.2702	0.4745	0.5152	0.5163	0.7173	0.8177
(1,0,0,1)	0.1021	0.3724	0.4042	0.6119	0.8176	0.8151
(1,0,0,0)	0.1042	0.1829	0.3433	0.4725	1.0000	0.6628
(0,1,1,1)	0.4486	0.1039	0.2873	0.1050	0.2873	0.2873
(0,1,1,0)	0.9575	0.5618	0.1025	1.0000	0.1046	0.1050
(0,1,0,1)	0.8894	0.9597	0.2915	0.5992	0.1049	0.2873
(0,1,0,0)	0.3915	1.0000	0.6306	1.0000	0.2873	0.9501
(0,0,1,1)	0.5098	0.7680	0.7680	0.7680	0.7680	0.5857
(0,0,1,0)	0.5293	0.5846	0.7680	0.5857	0.7680	1.0000
(0,0,0,1)	0.4382	1.0000	0.5832	0.5843	0.5853	1.0000
(0,0,0,0)	0.3701	0.4404	0.5722	0.5799	0.5856	0.7680

Table 7.14 Probability of Leaving the Job between a Fixed Period of Time

Personal Covariate	$\hat{\beta}_1 = 0.6348, \hat{\beta}_2 = 0.1289, \hat{\beta}_3 = 3.1846$						
	Up to 1 Yr	**Between 1 -2 Yrs**	**Between 2 – 3 Yrs**	**Between 3 – 4 Yrs**	**Between 4 – 5 Yrs**	**Between 5 – 6 Yrs**	**Total**
(1,1,1,1)	1.0000	-	-	-	-	-	1.0000
(1,1,1,0)	0.2312	-	-	0.7644	-	-	0.9956
(1,1,0,1)	0.1229	0.1243	-	0.2123	-	0.4512	0.9107
(1,1,0,0)	0.7454	0.1968	-	-	-	-	0.9422
(1,0,1,1)	0.8121	-	0.1385	-	-	-	0.9506
(1,0,1,0)	0.2325	0.1268	0.3175	0.2986	-	-	0.9754
(1,0,0,1)	0.6121	0.1125	0.0065	0.0231	0.0342	0.1276	0.9161
(1,0,0,0)	0.1895	0.2699	0.3256	0.0678	0.0981	0.0491	1.0000
(0,1,1,1)	-	-	1.0000	-	-	-	1.0000
(0,1,1,0)	0.5435	0.0764	0.1535	-	0.1543	-	0.9277
(0,1,0,1)	0.1129	0.2227	0.1821	0.0223	0.0245	0.3412	0.9057
(0,1,0,0)	0.7372	0.1968	-	-	-	-	0.9342
(0,0,1,1)	0.6436	0.1989	0.1085	-	-		0.9512
(0,0,1,0)	0.1525	0.2568	0.3975	0.0986	0.0596	0.0350	1.0000
(0,0,0,1)	-	-	-	1.0000	-	-	1.0000
(0,0,0,0)	0.2295	0.3999	0.1256	0.0678	0.0281	0.1291	0.9812

package, $u = 2.7439$, $b = 3.5742$. The hazard rate for the Extreme value distribution are given in table 7.15.

After the estimating the parameters of the Extreme value distribution the parameters $\beta_i, i = 1,2,3$ given in eqn. (2.6), eqn. (2.7) and eqn. (2.8) relating to the covariates are estimated as $\hat{\beta}_i = 1,2,3$ (0.5387, 0.2135, 2.8162).

The probability of leaving the job upto 6 years of service, given the personal covariates (0,0,0,0), (0,0,0,1), (0,0,1,0), (0,0,1,1), (0,1,0,0), (0,1,0,1), (0,1,1,0), (0,1,1,1), (1,0,0,0), (1,0,0,1), (1,0,1,0), (1,0,1,1), (1,1,0,0), (1,1,0,1), (1,1,1,0), ((1,1,1,1) are given in the following table 7.16, table 7.17 as per the above estimates.

RESULT AND DISCUSSION

The results from table 7.8 and table 7.9 shows that the Extreme value distribution is a better estimator of leaving pattern than the Weibull distribution. The chi-square value showed that Extreme value distribution could be used as a predictor for the company considered,

Table 7.16 Probability of Leaving a Job within 6 Years of Service

Personal Covariate	Up to 1 Yr	Between 1 -2 Yrs	Between 2 – 3 Yrs	Between 3 – 4 Yrs	Between 4 – 5 Yrs	Between 5 – 6 Yrs
(1,1,1,1)	0.9577	1.0000	0.9677	1.0000	1.0000	1.0000
(1,1,1,0)	1.0000	1.0000	0.0641	0.1572	0.3135	0.4322
(1,1,0,1)	0.4918	0.2347	0.7354	0.5677	0.5677	0.5677
(1,1,0,0)	0.5113	0.5666	1.0000	1.0000	1.0000	1.0000
(1,0,1,1)	0.4202	0.5245	0.5652	0.5663	0.5673	0.5677
(1,0,1,0)	0.3521	0.42245	0.5542	0.5619	0.7676	0.8677
(1,0,0,1)	0.2349	0.5431	0.5933	0.6225	0.7526	0.8128
(1,0,0,0)	0.7568	0.7768	0.4468	0.1768	0.3768	1.0000
(0,1,1,1)	0.8768	0.1091	0.8768	0.0902	0.8768	0.8768
(0,1,1,0)	0.0132	0.1420	0.3732	0.4663	0.6226	0.7413
(0,1,0,1)	0.8009	1.0000	1.0000	1.0000	1.0000	1.0000
(0,1,0,0)	0.8204	0.8757	0.6668	1.0000	1.0000	0.8768
(0,0,1,1)	0.7293	0.8336	0.8743	1.0000	0.8764	1.0000
(0,0,1,0)	0.6612	0.7317	0.8633	0.8710	0.8767	1.0000
(0,0,0,1)	0.1633	0.2420	0.4024	0.5316	0.6617	0.7219
(0,0,0,0)	0.4577	1.0000	0.5677	1.0000	0.5677	0.5677

Table 7.17 Probability of Leaving the Job between a Fixed Period of Time

Personal Covariate	$\hat{\beta}_1 = 0.5387, \hat{\beta}_2 = 0.2135, \hat{\beta}_3 = 2.8162$						
	Up to 1 Yr	Between 1 -2 Yrs	Between 2 – 3 Yrs	Between 3 – 4 Yrs	Between 4 – 5 Yrs	Between 5 – 6 Yrs	Total
(1,1,1,1)	1.0000	-	-	-	-	-	1.0000
(1,1,1,0)	0.6454	0.3322	-	-	-	-	0.9776
(1,1,0,1)	0.1334	0.0242	0.4235	0.1222	0.0145	0.2512	0.9690
(1,1,0,0)	0.5654	0.4346	-	-	-	-	1.0000
(1,0,1,1)	0.5324	0.0232	0.4143	-	-	-	0.9699
(1,0,1,0)	0.1435	0.2345	0.3678	0.0453	0.1768	0.0120	0.9799
(1,0,0,1)	0.6535	0.1353	-	-	0.0324	0.1324	0.9536
(1,0,0,0)	0.2773	0.1674	0.2753	0.1764	-	0.0343	0.9307
(0,1,1,1)	-	1.0000	-	-	-	-	1.0000
(0,1,1,0)	0.2235	0.1233	0.1278	0.1453	0.2268	0.1012	0.9479
(0,1,0,1)	0.3125	-	0.2742	0.1243	-	0.1569	0.8679
(0,1,0,0)	0.1273	0.4674	-	0.1232	0.1668	0.0242	0.9089
(0,0,1,1)	0.6535	-	0.0042	0.0243	0.1324	0.1249	0.9393
(0,0,1,0)	0.2473	-	0.3753	0.0764	-	0.2186	0.9176
(0,0,0,1)	0.7324	0.1232	0.1143	-	-	-	0.9699
(0,0,0,0)	0.5535	0.1145	0.1780	0.0453	0.0768	-	0.9681

with a reasonable degree of accuracy. From table 7.10 show that, the Chi-square value significant, the probability of leaving is dependent on the period of the stayers and leavers with respect to tenure, sex, qualification, location, etc.,. In the case of Cox's approach, the CLS distribution as Extreme value distribution, the probability of leaving the job is a sure event in the case of category of workers (1,1,1,1). In the case of the categories (1,1,0,1), (1,0,1,0) and (0,1,1,0) of persons, the probabilities are low but in all the cases the probabilities increase with the increase in the number of years of service. It is also seen that the category of workers under (1,0,1,0) have less probability of leaving in successive years but however the probabilities increase with the passage of time.

It is interesting to note that in both the CLS distributions namely Weibull distribution and Extreme value distributions, the probabilities are very low initially and approach unity as the years pass by for the group with covariates (1,0,1,0) and (1,0,0,1). It is quite interesting to observe that, for the category of persons with personal covariates as (1,1,1,1) the propensity to leave the job is very high throughout with the passage of time or CLS. This implies that the travel time is tedious, the overall incentives are satisfactory and the quality of life is not satisfactory is quite justified. Considering the combination of the personal covariates as (1,0,1,0) and (1,0,0,1) the probabilities are very low initially and slowly approach unity with the passage of time. From this one can understand that if a person's travel time is not tedious, the incentives are not satisfactory and the quality of life is not at all satisfactory, the propensity to leave the job is very much less initially but increases as the years pass by. Similar are the interpretations for all the combinations of personal covariates, the probability values depict the realities existing in practical life situations. The Extreme value distribution is better than the Weibull distribution particularly in estimation. However, this is not true in all cases and when predictions are being made for each company then they should examine to see which model is most appropriate.

Acknowledgment

The authors very much thankful to Mr. P. Rajesh, Programmer, Engineering Mathematics Section, Faculty of Engineering and Technology, Annamalai University for his valuable help in the computer simulation work.

Reference

Arivazhagan, G. R. Elangovan and R. Sathiyamoorthy. (2010). *A Stochastic Model for estimation of expected time to stop enrollment for recruitment*. International Journal of Physical Sciences, Vol. 22, No. 1 (M), pp. 277-284.

Arivazhagan, G., R. Elangovan, and R. Sathiyamoorthi (2010), *Determination of Optimal Manpower Reserve when Demand for Manpower has Fluctuation*. Vol. 29E, No.2, pp. 243-248.

Bartholomew, D. J., A. F. Forbes, and S. I. McClean (1991). *Statistical Techniques for Manpower Planning. Wiley,* Chichester.

Bartholomew, D. J. (1982). *The Stochastic Model for Social Processes*. 3rd ed., John Wiley and Sons, New York.

Bartholomew, D. J., A. F. Forbes, and S. I. McClean (1991). *Statistical Techniques for Manpower Planning*. Wiley, Chichester.

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

Biswas, S. (1988). *Stochastic Processes in Demography and Application.* Wiley Eastern Limited, Calcutta.

Biswas, S. (1996). *Statistical Techniques for Manpower Planning and Forecasting.* New age International Publisher, New Delhi.

Biswas, S. and S. K. Adhikari (1992). *Cox's Regression Model in the Analysis and Forecasting of Manpower Wastage*. Manpower Journal, Vol. 28, No.2, pp. 1-7.

Cox, D. R. (1972). *Regression models and Life tables*. J.R.S.S. Vol. 24, pp. 187-220.

Cox, D. R. (1975). *Partial likelihood*. Biometrika, Vol. 62, pp. 269-296.

Edwards, J.S. (1983). *A Survey of Manpower Planning Models and their Application*. J. Opl. Res. Soc., Vol. 34, No. 11, pp. 1031-1040.

Elangovan, R., C. Anantharaj and R. Sathiyamoorthi (2008). *Optimisation Model for Hiring of Specialized Service in Manpower Planning*. International Journal of Business Insights and Transformation, Vol. 1, No. 1, pp. 46-50.

Elangovan, R., C. Anantharaj and R. Sathiyamoorthi (2006b). *Shock Model Approach to the Determination of Optimal Time interval between Recruitments*, Journal of Ultra Scientist of Physical Sciences, Vol. 18, No. 2 (M), pp. 233-238.

Elangovan, R., N. Vijayasankar and R. Sathiyamoorthi (2006d). *Application of Cox's Regression Model in the Analysis of Wastage of Manpower Data*, Journal of Science, Vol. 43, pp. 173-186.

Fang, J., and G. Li (2005). *Contemporary Multivariate Analysis and Design of Experiments*. World Scientific, New Jercy.

John S. Edwards (1983), *A Survey of Manpower Planning Models and their Application*. J. Opt. Res. Soc., Vol. 34, No.11, pp. 1031-1040.

Kalbfleisch, J.D and R. L. Prentice. (2002). *The Statistical Analysis of Failure Time Data*, John Wiley and Sons, New York.

Kalbfleisch, J.D., and R.L. Prentice. (1980). *The Statistical Analysis of Failure Time Data.* John Wiley and sons. New York.

Kaplan, E.L., and P. Meier. (1958). *Non-parametric Estimation from Incomplete Observations*. Journal of the American Statistical Association, Vol.53, pp.458-479.

Lane, K.F., and J.E. Andrew (1955). *A Method of Labour Turnover Analysis*, Journal of Royal Statistical Society, Vol. 118 (A), pp. 296-323.

Lawless, J. W, (1982). *Statistical models and methods for life time data*. John Wiley & Sons, New York.

Leemis, L. M. (1995). *Reliability Probabilistic Models and Statistical Methods*. Prentice-Hall, Englewood Cliffs, New Jersey.

McClean. S. I. and J. O. Gribbin. (1987). *Estimation for Incomplete Manpower Data.* Applied Stochastic Models and Data Analysis, Vol. 3, pp.337-341.

McClean. S. I. (1975). *A comparison of the lognormal and transition models of wastage*, the Statistician, Vol. 25, No. 4, pp. 281-293.

Seal, H. L. (1945), *The mathematics of a population composed of K stationary strata each recruited from the stratum* *below and supported at lowest level by a uniform number of annual entrants*. Biometrika, Vol. 33, pp. 226-230.

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.

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.

Turnbull B.W. (1976). *The empirical distribution function with arbitrarily grouped, censored and truncated data.* JRSS, Vol. 38 (B), pp. 290-295.

Vijayasankar, N., Elangovan, R., and R. Sathiyamoorthy (2005). *Estimation of Expected time to recruitment using Shock Model Approach*. Bulletin of Pure and Applied Science, Vol. 24 E, No.1, pp.195-205.

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.

Young, A., and G. Almond (1961). *Predictions Distributions of Staff*. Computer Journal, Vol. 3, pp.246-250.

****

How to cite this article:

Arulpavai, R and Elangovan, R: Parametric and non-parametric estimation of incomplete manpower data using cox's approach. International Journal of Recent Scientific Research 2013; 4(3); 185-197.