CHAPTER III

PARAMETRIC AND NONPARAMETRIC MODELS
FOR SURVIVAL ANALYSIS

3.1 INTRODUCTION

The extensive accessibility of personal computers and rapid improvement of statistical computing have offered impetus to the development of complicated nonparametric and semi-parametric procedures in analyzing survival data. The traditional method of using a parametric approach in reliability investigations has been neglected by many biomedical researchers during the past four decades in favour of nonparametric procedures. Nonparametric methods require fewer assumptions than parametric methods. No distributional assumption is imposed on the survival times. In addition, there are many physical causes that led to the failure or death of an individual, and it is very difficult, if not impossible, to mathematically take into account all these factors into a specific functional form. In spite of these advantages and popularity of nonparametric and semi-parametric procedures, parametric modeling offers useful alternatives. Parametric models offer alternatives to the Cox model when the proportional hazards assumption is in question. If the distributional assumption on the survival times is valid, then the resulting estimates of the parameters, (e.g. the median) are more efficient in the sense of having smaller standard errors as compared to those in the nonparametric models. Interpretability of the results is
easier. In addition, the full likelihood is utilized in the analysis, resulting in more precise statistical inference.

A number of theoretical distributions have been used to approximate survival data. Fully parametric methods assume the knowledge of the distributions of the survival times e.g. Log-logistic, Exponential, Weibull, and Gompertz. Two parametric regression models, the exponential and Weibull were extensively used to compare and contract the analysis of right censored survival time data with usual linear regression model with normally distributed errors.

A convenient and clinically plausible way to describe survival time, shown here for a model with a single covariate, with the equation

\[ t = e^{(\beta_0 + \beta_1 x)} \varepsilon \quad ? (3.1.1) \]

where \( t \) express survival time, \( e^{(\beta_0 + \beta_1 x)} \) the product of the systematic component of the model, \( \varepsilon \) is the error component. This model can be linearized by taking the natural log of both side of the equation, yielding

\[ \ln(t) = \beta_0 + \beta_1 x + \varepsilon \quad ? (3.1.2) \]

where \( \varepsilon = \ln(\varepsilon) \). The error component, \( \varepsilon \) follows the extreme minimum value distribution, denoted \( G(0, ) \). Noted that under this model survival time followed the exponential distribution when \( =1 \) and the Weibull distribution when \( 1. \)

Non-parametric methods make no assumptions of the survival times. They are also called distribution-free methods the methods
are quiet easy to understand and apply. Kaplan and Mei (1958) product limit method is the most commonly used technique for estimating the survivorship function for samples of small and even for large sizes. The life table method is the next frequently used technique for estimating the survivorship function. The product limit estimates and life-table estimates of the survivorship function are essentially same. The only difference is that the product limit estimate is based on individual survival times while in the life-tables method survival times are grouped into intervals. The product limit estimate can be considered as a special case of the life-table estimate where each interval contains only one observation. However, if the data have been already grouped or large, it may be more convenient to perform a life-table analysis.

There are some other methods for comparing the survival functions like Log-rank, Tarone-Ware, Breslow etc., and these are less efficient than parametric methods when survival times follow a theoretical distribution and more efficient when no suitable theoretical distributions are known.

3.2 EXPONENTIAL MODEL

Exponential model was developed in details in both a homogeneous and a heterogeneous population setting. Researchers began to choose the exponential distribution to describe the life pattern of electronic systems even in the late 1940s. Bank statement and ledger error, payroll check errors, automatic calculating machine failure, and radar set component failure, in which the failure data
were well described by the exponential distribution of \( s \) (1952). Epstein and Sobel (1953) reported the reasons for selecting the exponential distribution over the population normal distribution and also showed how to estimate the parameter for singly censored data. Epstein (1958) made further discussion on the justification for the assumption of an exponential distribution. It has its futures like is often referred to as a purely random failure pattern and also it is famous for its unique lack of memory?, which requires that the age of animal or human does not affect future survival.

is the only parameter of constant hazard rate, when the survival time \( T \) follows Exponential distribution with parameter , the probability distribution function is defined as

\[
f(t) = \begin{cases} \lambda e^{\lambda t} & t \geq 0 \\ 0 & t < 0 \end{cases}
\]

The cumulative distribution function is

\[
F(t) = 1 - e^{\lambda t} \quad t \geq 0
\]

and the survivorship function and hazard function respectively

\[
S(t) = e^{-\lambda t} \quad t \geq 0
\]

\[
h(t) = \frac{f(t)}{S(t)} = \lambda \quad t \geq 0
\]

Hazard function is a constant and independent of \( t \). then consider the single covariate model in (3.1.2). The error distribution is log-exponential, that is, the extreme minimum value distribution denoted \( G(0,1) \). The survivorship function expressing in terms of time is

\[
S(t) = e^{-\lambda t} \quad t \geq 0
\]
\[ S(t, x, \beta) = \exp\left( \frac{t}{e^{\beta \cdot x}} \right) \] \hspace{1cm} (3.2.1)

To obtain the median survival time and set the right-hand side of (3.2.1) equal to 0.5 and solve the resulting equation, yielding the equation

\[ t_{50}(x, \beta) = e^{\beta \cdot x} \cdot \ln(0.5) \] \hspace{1cm} (3.2.2)

from the covariates in (3.2.2) is dichotomous, coded 0/1, the ratio of the median survival time for the group with \( x=1 \) to the group with \( x=0 \) denoted \( (x=1, x=0) \), is

\[
\frac{t_{50}(x = 1, \beta)}{t_{50}(x = 0, \beta)} = \frac{e^{\beta \cdot 1} \cdot \ln(0.5)}{e^{\beta \cdot 0} \cdot \ln(0.5)} = e^{\beta} \] \hspace{1cm} (3.2.3)

Alternatively, the relationship between the two median times is

\[ t_{50}(x = 1, \beta) = e^{\beta} \cdot t_{50}(x = 0, \beta) \] \hspace{1cm} (3.2.4)

An alternative way to present multiplicative effect \( \beta \) via the survivorship function. It follows from (3.2.3) that the relationship between the survivorship functions for the two groups is

\[ S(t, x = 1, \beta) = S(t, e^{\beta} \cdot x = 0, \beta) \] \hspace{1cm} (3.2.5)

The interpretation of the result in (3.2.5) is that the value of the survivorship function at time \( t \) for the group with \( x=1 \) may be obtained by evaluating the survivorship function for the group \( x=0 \) at time \( t \exp (-1) \). The change in the sign of the coefficient is due to the act that time percentiles and survival probabilities are inverse operations of one another.
The hazard function for the model is
\[ h(t; x, \beta) = e^{\beta' x} \]  \hfill (3.2.6)

And the hazard ratio for the dichotomous covariate is
\[ (x_1, x_0) = e^k \]  \hfill (3.2.7)

It assumes that the observations are subject to right censoring, but the analysis may be extended to other types of censoring and truncation using the methods for the proportional hazards model. Under the assumption that we have \( n \) independent observations of time, \( p \) covariates and a censoring indicator denoted \((t_i, x_i, c_i), i=1,2,\dotsc,n\), the log-likelihood function is
\[ L(\beta) = \prod_{i=1}^{n} c_{z_i} e^{z_i} \]  \hfill (3.2.8)

where, \( z_i = x_i' \beta, y_i = \ln(t_i), x_i = x_{i0}, x_{i1}, \dotsc, x_{ip} \) and \( x_{i0} = 1 \).

The likelihood equations are obtained by differentiating the log-likelihood function with respect to the unknown parameters and setting the expression equal to zero.

The general form of the second derivative of the log-likelihood function is
\[ \frac{\partial^2 L(\beta)}{\partial \beta_j \partial \beta_k} = \sum_{i=1}^{n} x_{ij} x_{ik} e^{x_i}, \quad j, k = 1, \dotsc, p \]  \hfill (3.2.9)

The estimators are based on the observed and observed information matrix, denoted \( I(\hat{\beta}) \), which is the matrix with elements given by the negative of (3.2.9) evaluated at the estimates of the coefficients. The inverse of the observed information matrix provides the estimators of the variances and covariance, namely
\[ \text{Var}(\hat{\theta}) \text{ I}(\hat{\theta})^1 \quad ? (3.2.10) \]

The end points of a 100(1- \alpha) percent Wald-statistic based confidence interval for the \( j \)-th coefficients are
\[ \hat{\beta}_j \pm z_{\alpha/2} \text{Se}(\hat{\beta}_j) \quad ? (3.2.11) \]

where \( \text{Se}(\hat{\beta}_j) \) denotes the estimator of the standard error of the estimator of the coefficient.

The estimator of the baseline hazard for the exponential model is obtained from (3.2.11) and is
\[ h_0(t, \hat{\beta}) = \exp(\hat{\beta}_0) \quad ? (3.2.12) \]

3.3 WEIBULL MODEL

The Weibull distribution is a generalization of the exponential distributions. It does not assume a constant hazard rate and has broader application. The applicability of the distribution to various failure situations were proposed by Weibull (1939, 1951). Moreover it has then been used in many studies of reliability and human disease mortality.

The basic form of a Weibull model was presented in
\[ \ln(T) = \beta_0 + \beta_1 x \quad ? (3.3.1) \]

The main difference between the Weibull and the exponential regression model is that the parameter in the distribution of the error term?. The inclusion of this parameter in the model leads to a slightly more complicated hazard function and related regression model parameters. For this reason the Weibull model in the single covariate setting to compare and contrast it with the potential
model. Thus, the Weibull distribution may be used to model the survival distribution of a population with increasing, decreasing, or constant risk.

The Hazard function for the single covariate Weibull regression model is

$$h(t, x, \beta, \lambda) = \frac{t^{-1}}{e^{\xi_x}} \quad ? (3.3.2)$$

where, for covariance, use \( =1/ \). This hazard function may be re-expressed in proportional hazards or accelerated failure time form. The proportional hazards form of the function is obtaines:

$$h(t, x, \beta, \lambda) \propto \lambda \left( \frac{t^{-1}}{e^{\xi_x}} \right) \quad \text{so}$$

$$h(t, x, \beta, \lambda) \propto \lambda \left( \frac{t^{-1}}{e^{\xi_x}} \right) \propto \lambda \left( \frac{t^{-1}}{e^{\xi_x}} \right)$$

$$h(t, x, \beta, \lambda) \propto \lambda \left( \frac{t^{-1}}{e^{\xi_x}} \right) \propto \lambda \left( \frac{t^{-1}}{e^{\xi_x}} \right)$$

where \( \gamma \exp(\beta_0/\sigma) \exp(\beta_1) \beta_1/\sigma \) and the baseline hazard function is \( h_0(t) \propto t^{-1} \). Although the parameter is a variance-like parameter on the log-time scale, \( =1/ \) is commonly called the shape parameter and \( =1/ \) is called the scale parameter, then (3.3.3) can be re-expressed as

$$h(t, x, \beta, \lambda) \propto \lambda \left( \frac{t^{-1}}{e^{\xi_x}} \right) \propto \lambda \left( \frac{t^{-1}}{e^{\xi_x}} \right) \propto \lambda \left( \frac{t^{-1}}{e^{\xi_x}} \right)$$

The survivorship function corresponding to the accelerated failure time form of the hazard function in (3.3.4) is

$$S(t, x, \beta, \sigma) = \exp \left( -t^\beta \exp\left( \frac{1/\sigma}{\beta_0 \beta_2 x} \right) \right) \quad ? (3.3.5)$$
The equation for the median survival time by setting the survivorship function equal to 0.5 and solving for timing

\[ t_{50}(x, \beta, \sigma) = \frac{\ln(0.5)}{e^{\beta} x} \quad (3.3.6) \]

If the covariate is dichotomous and coded 0/1, the timing ratio at the median survival time is like in the form given below

\[ (x \mid x \equiv 0) \frac{t_{50}(x \mid 1, \sigma)}{t_{50}(x \mid 0, \sigma)} \]

\[ \frac{[\ln(0.5)] e^{\beta} x}{[\ln(0.5)] e^{\beta}} \]

\[ e^{\beta} \quad (3.3.7) \]

The equation for the log-likelihood function for a singly containing right-censored data is obtained using

\[ z_i \quad y_i \quad x_i \beta \quad \bar{\sigma} \]

And \( f(z) \) is replaced with \( f(z)/f(z) \). This yields the log-likelihood function

\[ L(\beta, \sigma) = \prod_{i=1}^{n} c_i \left( \ln(\sigma) \right) e^{z_i} \quad (3.3.8) \]

The score equation for the \( j \)th regression coefficient is obtained by taking the derivative of (3.3.8) with respect to \( j \) and setting it equal to zero, yielding

\[ \frac{L(\beta, \sigma)}{\beta_j} = \prod_{i=1}^{n} x_{ij} c_i e^{z_i} \quad 0, \quad j = 1, \ldots, p \quad (3.3.9) \]

The score equation for the shape parameter, is

\[ \frac{L(\beta, \sigma)}{\sigma} = \prod_{i=1}^{m} z_i c_i e^{z_i} \quad 0 \quad (3.3.10) \]

The individual elements of this matrix to be evaluated are
\[ \frac{2L(\beta, \sigma)}{\beta_j \sigma_k} \frac{1}{\sigma_i^2} x_{ij} x_{ik} e_i^2 \]
\[ \frac{2L(\beta, \sigma)}{\beta_j} \frac{1}{\sigma_i^2} x_{ij} z_i e_i, \text{ and} \]
\[ \frac{2L(\beta, \sigma)}{\sigma \sigma} \frac{m}{\sigma_i^2} \frac{1}{\sigma_i^2} z_i^2 e_i. \]

when evaluated at the solution to the likelihood equations, the information matrix may be expressed as

\[ I(\beta, \sigma) = \frac{1}{\sigma^2} X' \Sigma X, \quad x' \Sigma x = m \]  \hspace{1cm} (3.3.11)

where \( X \) is an \( n \) by \( p+1 \) matrix containing the values of the covariates, \( \Sigma = \text{diag}(e_i^2) \), an \( n \) by \( n \) diagonal matrix, and

\[ z = (z_1, z_2, \ldots, z_n) \text{, with } z_i \frac{Y_i}{x_i} \frac{x_i' \beta}{\sigma}. \]

The estimator of the covariance matrix of the estimators of the parameters is

\[ \text{Var}(\beta, \sigma) = I(\beta, \sigma)^{-1} \]  \hspace{1cm} (3.3.12)

The details of the model building process for the Weibull regression model are the same as those presented in the above section for the exponential regression model. The martingale residuals used in the (Grambsch et al., 1995) plots for checking the scale of continuous covariates and for model assessment are

\[ \tilde{M}_i, \quad \exp(\tilde{z}_i) \]
\[ c_i t_i^2 \exp(\tilde{z}_i x_i \beta) \]  \hspace{1cm} (3.3.13)
where $\hat{\lambda} = 1/\hat{\varphi}$, the estimator of the cumulative hazard function for the Weibull regression model is

$$\hat{r}(t, x_i \hat{\beta}, \hat{\varphi}) = \exp(y_i x_i \hat{\beta}) \exp(\hat{Z}_i) \left(t, e^{x_i \hat{\beta}}\right)^{\hat{\lambda}}$$

for $i=1,2,\ldots,n$. Alternatively, one may calculate the values from the martingale residuals as

$$\hat{r}(t, x_i \hat{\beta}, \hat{\varphi}) = c_i \hat{M}_i$$

The estimator of the baseline hazard for the Weibull model is obtained from (3.3.13)

$$h_0(t, \hat{\lambda}, \hat{\varphi}) = \frac{1}{\hat{\varphi}^{\hat{\lambda}}} \exp(\hat{\beta}_0 / \hat{\varphi}) t^{(1/\hat{\lambda})}$$

(3.3.14)

3.4 GAMMA MODEL

The gamma distribution is a generalization of the exponential model. The gamma distribution with the parameters $\alpha > 0$ and $\beta > 0$ has the density function

$$f(t) = \frac{2^\alpha}{
\Gamma(\alpha)} t^{\alpha-1} e^{-t} \text{ for } \alpha > 0 \text{ and } \beta > 0 \quad (3.4.1)$$

for $t > 0$, where $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$.

Then,

$$E(T) = \frac{\alpha}{\beta}$$

and

$$\text{var}(T) = \frac{\alpha}{\beta^2}$$

The survival function is

$$S(t) = \exp \left( -t \int_0^t f(u) du \right) \quad (3.4.2)$$
The gamma model does not have closed form expression for $S(t)$ and $\lambda(t)$.

3.5 LOG-LOGISTIC MODEL

The log-logistic distribution has a non-monotonic hazard function, which makes it suitable for modeling survival data. A log-logistic regression model is described in which he hazard functions for separate samples converge with time. This also provides a linear model for the log odds on survival by any chosen time.

The survival time $T$ has a log-logistic distribution if $\log(t)$ has a logistic distribution. The density, survivorship, hazard and cumulative hazard functions of the log-logistic distribution are, respectively,

$$
\begin{align*}
\phi(t) &= \frac{\alpha \gamma^\alpha t^{\alpha-1}}{\left(1 + \gamma t\right)^{\alpha+1/2}} \\
S(t) &= \frac{1}{1 + \gamma t} \\
h(t) &= \frac{\alpha \gamma^\alpha t^{\alpha-1}}{\left(1 + \gamma t\right)^{\alpha+1}} \\
H(t) &= \log(1 + \gamma t), \ t \geq 0, \alpha > 0, \gamma > 0
\end{align*}
$$

The single covariate log-logistic accelerated failure time model may be expressed as

$$
\ln(T) = \gamma \ln(t) + \alpha \times \chi \\
\ln(T) = \gamma \ln(t) + \alpha \times \chi \\
\ln(T) = \gamma \ln(t) + \alpha \times \chi
$$

where the error term follows the standard logistic distribution. The appealing feature of the log-logistic model is that the slope coefficient can be expressed such a way that it can be interpreted as
an odds-ratio. In order develop the odds-ratio, by exp
   ng the
survivorship function for the model as
\[ S(t,x, , 1 \exp z) \quad ? (3.5.3) \]
where \( z = (y / \sigma + x)^{1/2} \). The odds of a survival time of at
least \( t \) is
\[ \frac{S(t,x, , 1 \exp z)}{\exp z} \quad ? (3.5.4) \]
It assumes that the covariate is dichotomous and coded 0/1.
The odds-ratio at time \( t \) formed from the ratio of the odds in (3.5.4)
evaluated at \( x=1 \) and \( x=0 \) is
\[ \text{OR}(t,x) \quad \frac{\exp \left(\frac{y}{0}^{1/0} \right)}{\exp \left(\frac{y}{0}^{1/0} \right)} \quad ? (3.5.5) \]
An alternative interpretation is obtained when expres the
median survival time as a function of the regression coefficients.
Setting the survivorship function in (3.5.5) equal to and solving,
obtain an equation for the median survival time of
\[ t_{50}(x, ) \quad \exp \left(\frac{0}{1} \right) \quad ? (3.5.6) \]
and the time ratio at the median is
\[ t_{50} \quad \exp \beta \quad ? (3.5.7) \]
As expected with an accelerated failure time model, the
exponentiated coefficient provides the acceleration factor on the time
scale. The percentile of the survival time distribution in (3.5.6) is of
the form

44
t_i(x, , ) 1 p / p exp 0 x

Cox and Oakes (1984) showed that the log-logistic mode is the only accelerated failure time model with the proportion odds property in (3.5.5). As in the exponential and Weibull models, maximum likelihood is the method usually employed to fit a log-logistic model to a set of data subject to right censoring. It follows from results for the standard logistic distribution (Evans et al., 1993; Klein and Moeschberger, 1997) that the contribution of a non-censored time to the likelihood is

\[ \frac{1}{\prod_i \exp z_i} \]  

and the censored time is

\[ 1 \prod_i \exp z_i \]

where for a multivariate model, \( z = \frac{y \times \beta}{\sigma} \). It follows that the log-likelihood function for a sample of \( n \) independent observations of time covariates and censoring indicator, denoted \((t_i, x_i, c_i), i = 1, 2, \ldots, n, \) is

\[
L = \prod_{i=1}^{n} c_i \ln z_i - 2 \ln \prod_i \exp z_i - 1 \ln 1 - \ln \exp z_i \quad (3.5.8)
\]

The score equation for the \( j \)th regression coefficient is obtained by taking the derivative of the log-likelihood in (3.5.8) with respect to \( \beta_j \), and the score equation for \( j \) is obtained in a similar manner. The estimator the multivariable form of the survivorship function in (3.5.5), and for the \( i \)th subject it is

\[
\hat{H}(t_i, \hat{\beta}) \ln 1 - \exp \hat{z}_i \quad \ldots \quad (3.5.9)
\]
where \( \hat{z}, (y_1, x_1) / \hat{\beta} \) and \( \hat{\beta} \) denote the estimators maximizing (3.5.8)

### 3.6 LOG-NORMAL MODEL

The lognormal distribution can be defined as the distribution of a variable whose logarithm follows the normal distribution. It is originated from McAlister (1879) and he described explicitly a theory of the distribution. Gaddum (1945) gave a review of its application in biology, followed by Boag (1949) application of cancer research. The properties, estimation problems as well as uses in economics have been discussed in detail by Aitchison and Brown (1957). Later other researchers such as Osgood (1958), Feinleib and Macmohan (1960) have observed that the distribution of age at onset of Alzheimer's disease followed the log normal distribution.

Consider the survival time \( T \) such that \( \log T \) is normally distributed with mean \( \mu \) and \( \sigma^2 \). \( T \) is lognormally distributed and can be rewritten as \( (, 2^*) \) where \( , \) \( \sigma^2 \) are two parameters of lognormal distribution.

The hazard function increases initially to a maximum and then decreases to zero as time approaches infinity (Watson Wells, 1961). Therefore, the lognormal distribution is suitable for survival patterns with an initially increasing and the decreasing hazard rates.

The popularity of lognormal distribution is to the fact that the cumulative values of \( y \log \) can be obtained from the tables of the
standard normal distribution and the corresponding values of \( t \) are then found by taking antilog. The probability density function and survivorship functions are respectively

\[
f_{t} = \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} \log_e t \right)^2 t > 0 \quad (3.6.1)
\]

and

\[
S_{t} = \frac{1}{\sqrt{2}}, \quad \frac{1}{x} \exp \left( -\frac{1}{2} \log_e x \right)^2 dx \quad (3.6.2)
\]

Let \( a = \exp \mu \). Then \( \mu = \log a \), (3.6.1) and (3.6.2) can be written as

\[
f_{t} = \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} \log_e \log at \right)^2 \quad (3.6.3)
\]

and

\[
S_{t} = \frac{1}{\sqrt{2}}, \quad \frac{1}{x} \exp \left( -\frac{1}{2} \log_e \log at \right)^2 \frac{dx}{x} \quad (3.6.4)
\]

\[1 \ G \log _e \log at / \quad (3.6.5)\]

where \( G(y) \) is the cumulative distribution function of \( \log \) standard variable

\[
G_y = \frac{1}{\sqrt{2}} \int_{0}^{y} e^{-u^2/2} du \quad (3.6.6)
\]

The lognormal distribution is completely specified by the two parameters \( \mu \) and \( \sigma^2 \). Time \( T \) cannot assume zero value since \( \log_e T \) is not defined for \( T = 0 \). The hazard function, from (3.6.3) and (3.6.5), has the form

\[
h_{t} = \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} \log_e \log at \right)^2 \quad (3.6.7)
\]

\[
\frac{1}{G \log_e \log at /} \quad (3.6.7)
\]
The mean and variance of two parameter lognormal distribution are respectively

\[ \exp \left( \frac{1}{2} \mu^2 \right) \text{ and } \exp \left( \mu^2 + \sigma^2 \right) \]

The coefficient of variation of the distribution is \( \exp \left( \frac{1}{2} \sigma^2 \right) - 1 \). The median is \( e^{\mu} \) and the mode is \( \exp \left( \mu - \frac{\sigma^2}{2} \right) \).

3.7 LIFE TABLE METHOD

The simplest methods of analyzing survival data is the life table method and are particularly suited for analyzing very data sets, Young et al., (1999). Life tables are different from other methods of survival analysis in which observations are grouped into discrete time intervals. The life table method is one of the oldest methods for measuring mortality and describing the survival experience of a population. It has been used by actuaries, demographers, governmental agencies, and medical researchers in various studies like survival, population growth, fertility, migration, and so on. In attendance it has been a decennial series of life tables on the entire U.S. population since 1900. The life-table provides a summary, for a population or sub-population, of the relationship of mortality to age, based on prevailing mortality rates. It includes for each age \( x \), measures such as life expectancy at age \( x \) and the probability of dying before age \( x+1 \). Life-tables are used by demographers interested in comparing detailed mortality schedules between countries and sub-populations, and by actuaries in the calculation of life-insurance premiums. The life tables are categorized into two types. One is
population life tables and another is clinical life tables. When the life tables are summarizes the mortality experience of a specific population for a period of time are called population tables. When the life tables have been applied to patients with a given disease who have been followed for a period of time and constructed for patients are called clinical life tables. Moreover the calculation of population and clinical life tables are similar only the difference is in the source of data, which are required.

The life table method assumes that subjects are withdrawn randomly throughout each interval and an average they withdrawn half way through out the interval. This is not an important issue when the time intervals are short but bias may introduced when time intervals are long. This method also assumes that the rate of failure within an interval is same for all subjects and is independent of the probability of survival at other time periods. Lif tables are produced from large scale population surveys and are less-frequently used these days because of the reason the Kaplan-Meier method being preferred because it is less prone to bias.

The clinical life table method is nothing but actuarial life table method and it has been applied to clinical data for many decades. Berkson and Gage (1950) and Cutler and Ederer (1958) gave a life-table method for estimating the survivorship function. The life-table method requires a reasonably large number of observations so that survival times can be grouped into intervals and it incorporates all survival information accumulated up to the termination of the study.
3.8 KAPLAN MEIER ESTIMATOR

The Kaplan and Meier or product limit estimator provides a non-parametric maximum likelihood estimate of the survivor function (Kaplan and Meier, 1958) and it is a special case of the life table technique in that series of time intervals is formed such a way that only one death occurs in each interval and the death occurs at the beginning of the interval. It estimates the probability of surviving longer than a given time \( t \) and it is the product of a series of estimated conditional probabilities. The Kaplan Meier estimate of \( S(t) \) is given as

\[
S(t) = \prod_{i=1}^{t} \left( \frac{n_i}{i} \right)^{c_i} \quad (3.8.1)
\]

\( S(t) \) is estimated survival function at time \( t \), \( \prod_{i=1}^{t} \) denotes the multiplication of the survival times across all cases \( i \) than or equal to \( t \) (the geometric mean), \( n = \) total number of cases in the sample and \( i \) is the number of cases surviving up to time \( t \). \( C_i \) is a constant such that the code 1 is uncensored case or terminal case and 0 is coded for censored cases or it can be derived in the form as

\[
S(t) = \prod_{i=t}^{n} \left( \frac{d_i}{n_i} \right) \quad (3.8.2)
\]

where \( n_i \) corresponds to the number of observations at risk of failing just prior to time \( t \) and \( d_i \) denotes the number of failures at time \( t \).

The Kaplan-Meier method is useful in estimating the survival distribution \( S(t) \). However, the product-limit estimates are limited to the time interval in which the observations fall. If the largest
observation is uncensored the product limit estimate at that time is always zero. If the largest observation is censored the product limit estimate can never equal to zero and is undefined beyond the largest observation unless an additional assumption is imposed. In adding together, if less than 50 percent of the observations are uncensored and the largest observation is censored then the median survival time cannot be estimated. Thus the method is not perfect and there are reasons to search for a parametric model.

$$S(t)$$ is computed at every distinct survival time and do not have to be concerned about the intervals between the distinct survival time in which no one dies and $$S(t)$$ remains constant and $$S(t)$$ is a step function starting with 1.0 and decreasing in steps of $$\frac{1}{n}$$ to zero. When $$S(t)$$ is plotted versus $$t$$ the various percentiles of survival time can be read from the graph or calculated from $$S(t)$$. The rule which may be generalized as the probability of surviving $$k$$ (2) or more years from the beginning of the study is a product of $$k$$ observed survival rates

$$S(k) = p_1 \times p_2 \times p_3 \times \ldots \times p_k$$

where $$p_1$$ denotes the proportion of patients surviving at least one year

$$p_2$$ denotes the proportion of patients surviving in the second year after they have survived one year,

$$p_3$$ denotes the proportion of patient surviving in the third year after they have survived two years, and
\( p_k \) denotes the proportion of patients surviving in the \( k^{th} \) year after they have survived \( k-1 \) years.

The product limit estimate of the probability of surviving any particular number of years from the beginning of the study is the product of the same estimate up to the previous year and the observed survival rate for the particular year:

\[
S(t) = S(t - 1)p_k, 
\]

To summarize this procedure, let \( n \) be the total number of individuals whose survival times are available. Then \( n \) survival times in order of increasing magnitude such that \( t_{(1)} \leq t_{(2)} \ldots \leq t_{(n)} \), then

\[
\hat{S}(t) = \prod_{r=1}^{n} \frac{n-r+1}{n} \quad (3.8.3)
\]

Where \( r \) runs through those positive integers for which \( t_{(r)} \leq t \) and \( t_{(r)} \) is uncensored. The estimated median survival time is the 50th percentile which is the value of \( t \) as \( \hat{S}(t) = 0.50 \).

The Kaplan-Meier method was also applied to health economics by Fenn et al., (1995) and it is used in the economic valuation of cost effectiveness of treatment where censored cost data are present. Censored data arises when the course of treatment extends beyond the end of the clinical trial period when patients withdraw from the trial for reasons unconnected with the treatment under study.

3.9 LOG RANK TEST

Researchers may wish to compare the survival curves of two different groups. The log rank test is a large sample \( \chi^2 \)-square test.
that uses its test criterion or stochastic that provides an over all comparison of the Kaplan and Meier being compared. The log rank statistic makes use of observed versus expected cell counts over categories of outcomes. The categories of log rank tests are defined by each of the ordered failure times for the entire set of data being analyzed. It is more powerful than the Breslow test if the mortality (number of terminated cases) of the groups is proportional i.e., the mortality of the groups differs by a constant multiplier. Life tables and the Kaplan-Meier product limit may be used to plot estimates of the survival curve for each of two groups of patients while the log-rank test is used to determine whether there is a statistically significant difference between the two survival curves Young et al., (1999). Use of the log-rank test is necessary when there are censored observations otherwise, nonparametric statistical methods such as the Mann-Whitney U test or Wilcoxon rank sum test may be used to compare survival times.

The log-rank test was developed by Peto and Peto (1972) based on the Savage (1956) test and its generalization of Mantel? (1966), is based on a set of scores \( w_i \), which are functions of logarithm of the several function Altshuler (1970) estimate the log survival function at \( t_{(i)} \) using

\[
e(t_{(i)}) = \frac{m_{(i)}}{r_{(i)}} \tag{3.9.1}
\]

where \( m_{(i)} \) is number of failure and \( r_{(i)} \) the number of observations, the score \( w_i = \ln \left( \frac{m_{(i)}}{r_{(i)}} \right) \) for uncensored observation \( t_{(i)} \) and \( -e(T) \) for an
observation censored at $T$. The log rank test is based on
the $w$ scores and the permutation variance of $S$ is

$$\text{Var}(S) = \frac{n_1 n_2 \sum_{i=1}^k w_i^2}{(n_1 \ n_2)(n_1 \ n_2 \ n_2)} \quad ? (3.9.2)$$

which can be rewritten as

$$V = \frac{k \ n_{ij} \ m_{ij}}{n_{ij} \ \frac{m_{ij}}{n_{ij} \ n_2 \ n_2}} \quad ? (3.9.3)$$

The statistic $S / \sqrt{\text{Var}(S)}$ has an asymptotically standard normal distribution under the null hypothesis. The log rank statistics $S$ can be shown to equal the sum of the observed failures minus the conditional expected failures computed at each failure time in one of the group. Let $O_1$ and $O_2$ be the observed numbers and $E_1$ and $E_2$ the expected numbers of death in the two treatment groups.

3.10 BRESLOW TEST

It is more powerful than the Log-rank test if the mortality of the groups is not proportional. The power of the Breslow Test declines as the number of censored cases increases. The Gehan-Breslow test is more powerful with data from the lognormal survival distribution but may have low power if there is heavy censoring. The Gehan-Breslow test gives more weight to earlier failures (deaths).

3.11 TARONE WARE TEST

Tarone and Ware (1977) with its intermediate weighting scheme, is designed to have good power across a wide range of survival
functions although it may not be the most powerful of the three tests in a particular situation. The statistic is of the form

$$ T = \frac{\sum_{i=1}^{k} w_i (d_i - E_i)}{\sqrt{\sum_{i=1}^{k} \left( \frac{w_i^2 v_i}{n_i} \right)}} $$  \hspace{1cm} (3.11.1) \\

The statistics was considered by (Cox, 1972; Peto and Peto; 1972) and is sometimes referred as logrank test. Under fairly general conditions (Tarone and Ware, 1977) has shown that $T$ is asymptotically normal with mean 0 and variance 1. Hence one may test the hypothesis of no difference by computing its value to the normal critical value. By the definition of Tarone and Ware procedure the weights assigned to individual event time are greater than the log rank weights and also Tarone and Ware procedure always superior to the least powerful statistics of logrank and Wilcoxon

The three important tests of non-parametric tests are log-Rank test (Mantel-Haenszel Test), Breslow (Generalized Wilcoxon test) and Tarone-Ware test. The equation is of the form

$$ U = \frac{W_i (D_i - E_i)}{\sum_{i=1}^{k} \left( \frac{w_i^2 v_i}{n_i} \right)} $$  \hspace{1cm} (3.11.2) \\

where $W_i$ denotes weight, $D_i$ the number of terminal events observed and $E_i$ is the number of terminal events expected which is calculated by the number at risk cases and termination at each event time $t$. The three statistical tests differ in the weighing factor and they use ($W_i$). When we look at log-rank test all the cases weight equally and this is considered as the least conservative among the three tests. Where as Breslow test is considered as the most conservative of the
other tests also \( E_i \) is the number of cases at risk at event time \( t \).
The Tarone-Ware test is considered as mid-conservative between the
other tests of log-rank and Breslow. Moreover the weighting factor
concern this weighs earlier cases but less than Breslow test does. \( W \)
specifies square root of the number of cases at risk at event time \( t \).

The statistical power of the three tests \((1 - \beta)\) varies from all the
three tests as discussed here. The log-rank test is more powerful than
the Breslow test only if the number of terminated case of the groups
is proportional i.e, the mortality of the groups by a constant
multiplier. Another important point of the power of Breslow test is that
it declines as the number of censored cases increases.

3.12 APPLICATION TO CLINICAL TRIAL FOR PARAMETRIC METHODS

The data base consists of 371 skin TB patients, who were
allocated for treatments in a clinical controlled trial of National
Institute for Research in Tuberculosis, Chennai. The covariates
considered here for the analysis are the different types of treatments,
gender, age, smear results for first five months period, the survival
time. The status for curing the disease has been obtained from the
smear test recorded for five months. If the smear test goes a certain
level then the disease is treated as cured otherwise it may require
some time to get cured.

Software function of survival analysis

Parametric Distribution Analysis is used to fit a chosen
distribution to the collected data and get estimates that describe the
survival probability (probability that the disease is red).

Based on the fitted distribution, one can then

Obtain parameter estimates and distribution characteristics such as Mean Time to Failure (MTTF).

Estimate percentiles and survival probabilities

Compare the fitted distribution to other distributions

Display probability, survival, cumulative failure, and hazard plots

The parameter estimates (shape and scale for Weibull and exponential; location and scale for smallest extreme value, normal, lognormal, logistic, log logistic; and additionally threshold for 3-parameter Weibull, 2-parameter exponential, 3-parameter lognormal, 3-parameter log logistic) define the best-fitting parameter estimates for the chosen distribution. One cannot determine from the estimated distribution parameters whether or not the chosen distribution fits the data well. Use the distribution plot, probability plot, and goodness-of-fit measures to determine if the distribution adequately fits the data.

**Goodness of fit**

Measure used to determine how well a particular distribution fits the data. Minitab provides Anderson-Darling and Pearson correlation coefficient goodness of fit statistics.

**Anderson-Darling (adjusted)**

The Anderson-Darling measures the area between the fitted line (based on chosen distribution) and the nonparametric survival function
(based on the plot points). More precisely, the Anderson-Darling statistic is a squared distance that is weighted more heavily in the tails of the distribution.

Smaller Anderson-Darling values indicate that the distribution fits the data better. Use the Anderson-Darling values to compare the fit of competing distributions as opposed to an absolute measure of how a particular distribution fits the data.

Minitab uses an adjusted Anderson-Darling statistic, because the statistic changes when a different plot point method is used.

**Data Analysis**

The survival times of patients who died from other causes, or lost to follow-up process are regarded as censored. A variable associated with the status of an individual at the end of the study takes the value 1 if the patient has free from skin TB and 0 if the survival time is right censored.

The details of analysis for fitting different distributions

**Exponential distribution**

Estimation Method: Maximum Likelihood

*Table 3.1 Distribution Analysis for time*

<table>
<thead>
<tr>
<th>Step</th>
<th>-Log Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>675.213</td>
</tr>
<tr>
<td>1</td>
<td>675.005</td>
</tr>
<tr>
<td>2</td>
<td>675.005</td>
</tr>
<tr>
<td>3</td>
<td>675.005</td>
</tr>
</tbody>
</table>

Log-Likelihood = -675.005
Table 3.2 Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>95.0% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Lower</td>
</tr>
<tr>
<td>Mean</td>
<td>3.71233</td>
<td>0.217248</td>
<td>3.31004</td>
</tr>
</tbody>
</table>

Goodness-of-Fit

Anderson-Darling (adjusted) = 182.206

Table 3.3 Characteristics of Distribution

<table>
<thead>
<tr>
<th>Descriptive Measures</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>95.0% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Lower</td>
</tr>
<tr>
<td>Mean (MTTF)</td>
<td>3.71233</td>
<td>0.217248</td>
<td>3.31004</td>
</tr>
<tr>
<td>Median</td>
<td>2.88592</td>
<td>2.57319</td>
<td>0.150585</td>
</tr>
<tr>
<td>First Quartile(Q1)</td>
<td>1.06797</td>
<td>0.0624982</td>
<td>0.952240</td>
</tr>
<tr>
<td>Third Quartile(Q3)</td>
<td>5.14638</td>
<td>0.301169</td>
<td>4.58869</td>
</tr>
<tr>
<td>Inter-quartile Range(IQR)</td>
<td>4.07841</td>
<td>0.238671</td>
<td>3.63645</td>
</tr>
</tbody>
</table>

Table 3.4 Survival Probability

<table>
<thead>
<tr>
<th>Time</th>
<th>Survival probability</th>
<th>95.0% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Lower</td>
</tr>
<tr>
<td>1</td>
<td>0.763859</td>
<td>0.739256</td>
</tr>
<tr>
<td>2</td>
<td>0.583480</td>
<td>0.546500</td>
</tr>
<tr>
<td>3</td>
<td>0.445696</td>
<td>0.404003</td>
</tr>
<tr>
<td>4</td>
<td>0.340449</td>
<td>0.298662</td>
</tr>
<tr>
<td>5</td>
<td>0.260055</td>
<td>0.220788</td>
</tr>
</tbody>
</table>
Figure 3.1 Estimated Survival Function, Hazard Function and Probability Distribution Function

An exponential distribution fitted to the survival time given in the data set of table 3.1. For these data, the total observed and right censored times is 1081 months, the number of uncensored observation is 292 and the number of censored observation is 66.

The estimated parameter of the distribution is \( \hat{\lambda} = \frac{292}{1081} = 0.27012 \). Therefore, \( \hat{\mu} = 3.7123 \) and the standard error of \( \hat{\mu} \) is 0.217248. The estimated hazard function is therefore \( \hat{\lambda}(t) = 0.27012, \ t > 0 \) and the estimated survivor function is \( \hat{S}(t) = \exp(-0.27012t) \).

The estimated median survival time is 2.885 months and an estimate of 75th percentile of the distribution of survival time is \( \hat{S}(75) = 5.1464 \). That is, 75% of Skin TB patient who are undergoing for regular treatment will have chances to get free from a disease of time less than 5 months.
The standard error of the estimated median time is 2.57. The
limit of a 95% confidence interval for the median survival time is from
0.150585 to 2.29435 (zero to 3 months approximately). The estimated
survival probabilities for different survival time are shown in
Table 3.4. The estimated hazard function, cumulative hazard function,
probability density function and the survivor function are shown in
figure 3.1.

**Weibull distribution**

Estimation Method: Maximum Likelihood

*Table 3.5 Distribution Analysis for time*

<table>
<thead>
<tr>
<th>Step</th>
<th>- Log Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>603.269</td>
</tr>
<tr>
<td>1</td>
<td>593.193</td>
</tr>
<tr>
<td>2</td>
<td>592.020</td>
</tr>
<tr>
<td>3</td>
<td>591.995</td>
</tr>
<tr>
<td>4</td>
<td>591.995</td>
</tr>
<tr>
<td>5</td>
<td>591.995</td>
</tr>
</tbody>
</table>

Log-Likelihood = -591.995

*Table 3.6 Parameter Estimates*

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>95.0% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shape</td>
<td>2.07112</td>
<td>0.100785</td>
<td>1.88271, 2.27838</td>
</tr>
<tr>
<td>Scale</td>
<td>3.69230</td>
<td>0.104949</td>
<td>3.49223, 3.90384</td>
</tr>
</tbody>
</table>

Goodness-of-Fit

Anderson-Darling (adjusted) = 166.861
Table 3.7 Characteristics of Distribution

<table>
<thead>
<tr>
<th>Descriptive Measures</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>95.0% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Lower</td>
</tr>
<tr>
<td>Mean(MTTF)</td>
<td>3.27062</td>
<td>0.0928063</td>
<td>3.09369</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>1.65674</td>
<td>0.0826625</td>
<td>1.50239</td>
</tr>
<tr>
<td>Median</td>
<td>3.09345</td>
<td>0.0946021</td>
<td>2.91348</td>
</tr>
<tr>
<td>First Quartile(Q1)</td>
<td>2.02322</td>
<td>0.0869173</td>
<td>1.85984</td>
</tr>
<tr>
<td>Third Quartile(Q3)</td>
<td>4.32304</td>
<td>0.123749</td>
<td>4.08718</td>
</tr>
<tr>
<td>Inter-quartile Range(IQR)</td>
<td>2.29982</td>
<td>0.107235</td>
<td>2.09896</td>
</tr>
</tbody>
</table>

Table 3.8 Survival Probabilities

<table>
<thead>
<tr>
<th>Time</th>
<th>Survival Probability</th>
<th>95.0% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Lower</td>
</tr>
<tr>
<td>1</td>
<td>0.935342</td>
<td>0.914226</td>
</tr>
<tr>
<td>2</td>
<td>0.755115</td>
<td>0.715356</td>
</tr>
<tr>
<td>3</td>
<td>0.521795</td>
<td>0.477946</td>
</tr>
<tr>
<td>4</td>
<td>0.307182</td>
<td>0.266115</td>
</tr>
<tr>
<td>5</td>
<td>0.153548</td>
<td>0.119875</td>
</tr>
</tbody>
</table>
Figure 3.2 *Estimated Survival Function, Hazard Function and Probability Distribution Function*

For comparison, a Weibull distribution will be fitted to the same survival data set. The distribution can be fitted using the computer software and from the resulting output, the estimated scale parameter of the distribution is found to be \( \hat{\lambda} = 3.69230 \) and the estimated shape parameter is \( \hat{\gamma} = 2.07112 \). The standard errors for these estimates are given by \( SE(\hat{\lambda}) = 0.104979 \) and \( SE(\hat{\gamma}) = 0.100785 \) respectively. Confidence interval for these estimates are \( (3.4922, 3.9038) \) and \( (1.88271, 2.278) \) respectively.

The estimated mean survival time of the distribution is \( \hat{\mu} = 3.27062 \) and the standard error of \( \hat{\mu} \) is 0.0928063. The estimated hazard function is therefore \( \hat{h}(t) = 7.6472t^{1.07112}, \ t > 0 \) and the estimated survivor function is \( \hat{S}(t) = \exp(-3.69230t^{2.07112}) \).

The estimated median survival time is 3.09345 months and an estimate of 75th percentile of the distribution of survival time is 7(75)
=4.32304. That is, 75% of Skin TB patient who were undergone for regular treatment will have chances to get free from the disease of time less than 4.3 months.

**Lognormal distribution**

Estimation Method: Maximum Likelihood

*Table 3.9 Distribution Analysis for time*

<table>
<thead>
<tr>
<th>Step</th>
<th>-Log Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>590.430</td>
</tr>
<tr>
<td>1</td>
<td>590.277</td>
</tr>
<tr>
<td>2</td>
<td>590.277</td>
</tr>
<tr>
<td>3</td>
<td>590.277</td>
</tr>
</tbody>
</table>

Log-Likelihood = -590.277

*Table 3.10 Parameter Estimates*

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>95.0% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Lower</td>
</tr>
<tr>
<td>Location</td>
<td>1.05658</td>
<td>0.0334367</td>
<td>0.991045</td>
</tr>
<tr>
<td>Scale</td>
<td>0.612285</td>
<td>0.0261460</td>
<td>0.563125</td>
</tr>
</tbody>
</table>

Goodness-of-Fits

Anderson-Darling (adjusted) = 167.657

*Table 3.11 Characteristics of Distribution*

<table>
<thead>
<tr>
<th>Descriptive Measures</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>95.0% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Lower</td>
</tr>
<tr>
<td>Mean(MTTF)</td>
<td>3.46955</td>
<td>0.134068</td>
<td>3.21649</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>2.33992</td>
<td>0.183274</td>
<td>2.00692</td>
</tr>
<tr>
<td>Median</td>
<td>2.87652</td>
<td>0.0961813</td>
<td>2.69405</td>
</tr>
<tr>
<td>First Quartile(Q1)</td>
<td>1.90332</td>
<td>0.0685751</td>
<td>1.77355</td>
</tr>
<tr>
<td>Third Quartile(Q3)</td>
<td>4.34732</td>
<td>0.171702</td>
<td>4.02349</td>
</tr>
<tr>
<td>Inter-quartile Range</td>
<td>2.44400</td>
<td>0.144323</td>
<td>2.17689</td>
</tr>
</tbody>
</table>
Table 3.12 Survival Probabilities

<table>
<thead>
<tr>
<th>Time</th>
<th>Survival probability</th>
<th>95.0% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Lower</td>
</tr>
<tr>
<td>1</td>
<td>0.957793</td>
<td>0.940108</td>
</tr>
<tr>
<td>2</td>
<td>0.723599</td>
<td>0.684628</td>
</tr>
<tr>
<td>3</td>
<td>0.472635</td>
<td>0.429962</td>
</tr>
<tr>
<td>4</td>
<td>0.295116</td>
<td>0.254898</td>
</tr>
<tr>
<td>5</td>
<td>0.183278</td>
<td>0.149019</td>
</tr>
</tbody>
</table>

Distribution Overview Plot for time

LSXY Estimates-Censoring Column in status

<table>
<thead>
<tr>
<th>Table of Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loc 1.07374</td>
</tr>
<tr>
<td>Scale 0.619630</td>
</tr>
<tr>
<td>Mean 3.54562</td>
</tr>
<tr>
<td>SDev 2.42573</td>
</tr>
<tr>
<td>Median 2.92631</td>
</tr>
<tr>
<td>IQR 2.51784</td>
</tr>
<tr>
<td>Failure 250</td>
</tr>
<tr>
<td>Censor 66</td>
</tr>
<tr>
<td>AD 0.0000</td>
</tr>
<tr>
<td>Correlation 0.953</td>
</tr>
</tbody>
</table>

Figure 3.3 Estimated Survival Function, Hazard Function and Probability Distribution Function

For comparison, a lognormal will be fitted to the same survival data set. The distribution can be fitted using the computer software, and from the resulting output, the estimated location parameter of the distribution is found to be 1.05658 and the estimated scale parameter is 0.612285. The standard errors for these estimates are given by
0.0334367 and 0.0261460 respectively. Confidence interval for these estimates are (0.991, 1.12211) and (0.563, 0.6657) respectively.

The estimated mean survival time of the distribution is \( \hat{\mu} = 3.46955 \) and the standard error of \( \hat{\mu} \) is 0.134068. The estimated median survival time is 2.87652 months and an estimate of 75th percentile of the distribution of survival time is \( \hat{\theta}(75) = 4.34732 \). That is, 75% of Skin TB patient who were undergone for regular treatment will have chances to get free from the disease of time less than 4.34 months.

The standard error of the estimated median time is 0.09618 months. The limit of a 95% confidence interval for the median survival time is from 2.7 to 3.1 months. The estimated survival probabilities for different survival time are shown in table 3.12. The estimated hazard function, cumulative hazard function, probability density function and the survivor function are shown in figure 3.3.

**Log logistic distribution**

Estimation Method: Maximum Likelihood

<table>
<thead>
<tr>
<th>Step</th>
<th>- Log Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>595.804</td>
</tr>
<tr>
<td>1</td>
<td>595.252</td>
</tr>
<tr>
<td>2</td>
<td>595.247</td>
</tr>
<tr>
<td>3</td>
<td>595.247</td>
</tr>
</tbody>
</table>

Log-Likelihood = -595.247
Table 3.14 Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>95.0% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td>1.07561</td>
<td>0.0338374</td>
<td>1.00929</td>
</tr>
<tr>
<td>Scale</td>
<td>0.361642</td>
<td>0.0174619</td>
<td>0.328986</td>
</tr>
</tbody>
</table>

Goodness-of-Fit

Anderson-Darling (adjusted) = 167.115

Table 3.15 Characteristics of Distribution

<table>
<thead>
<tr>
<th>Descriptive Measures</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>95.0% Normal CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (MTTF)</td>
<td>3.67238</td>
<td>0.151075</td>
<td>3.38791</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>3.47578</td>
<td>0.447727</td>
<td>2.70027</td>
</tr>
<tr>
<td>Median</td>
<td>2.93179</td>
<td>0.0992041</td>
<td>2.74366</td>
</tr>
<tr>
<td>First Quartile (Q1)</td>
<td>1.97054</td>
<td>0.0760753</td>
<td>1.82694</td>
</tr>
<tr>
<td>Third Quartile (Q3)</td>
<td>4.36194</td>
<td>0.170927</td>
<td>4.03947</td>
</tr>
<tr>
<td>Inter-quartile Range</td>
<td>2.39140</td>
<td>0.147129</td>
<td>2.11973</td>
</tr>
</tbody>
</table>

Table 3.16 Survival Probabilities

<table>
<thead>
<tr>
<th>Time</th>
<th>Survival Probabilities</th>
<th>95.0% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Lower</td>
</tr>
<tr>
<td>1</td>
<td>0.951397</td>
<td>0.933453</td>
</tr>
<tr>
<td>2</td>
<td>0.742228</td>
<td>0.700607</td>
</tr>
<tr>
<td>3</td>
<td>0.484105</td>
<td>0.438515</td>
</tr>
<tr>
<td>4</td>
<td>0.297530</td>
<td>0.257123</td>
</tr>
<tr>
<td>5</td>
<td>0.186014</td>
<td>0.153349</td>
</tr>
</tbody>
</table>
Figure 3.4 Estimated Survival Function, Hazard Function and Probability Distribution Function

For comparison, a log logistic distribution will be fitted to the same survival data set. The distribution can be fitted using the computer software and from the resulting output the estimated scale parameter of the distribution is found to be 0.3616 and the estimated location parameter is $\hat{\theta} = 1.07562$. The standard errors for these estimates are given by $SE(\theta) = 0.0174619$ and $SE(\hat{\theta}) = 0.0338374$ respectively. Confidence interval for these estimates are (1.00929, 1.4193) and (0.328986, 0.397538) respectively.

The estimated mean survival time of the distribution is 3.67238 and the standard error of estimated mean survival time is 0.151075. The estimated hazard function is therefore $\lambda(t) = 1.0602t^{0.638/}}$
$2.9317t^{0.36164}$, $t>0$ and the estimated survivor function is $\hat{S}(t) = e^{\exp \left(1 + 2.9317t^{0.36164}\right)}$.

The estimated median survival time is 2.93179 months an estimate of 75th percentile of the distribution of survival time is $\hat{2}(75) = 4.36194$. That is, 75% of Skin TB patient who were undergone regular treatment will have chances to get free from the disease of time less than 4.36 months.

The standard error of the estimated median time is 0.099 months. The limit of a 95% confidence interval for the median survival time is from 2.7 to 3.1 months. The estimated survival probabilities for different survival time are shown in table 3.16. The estimated hazard function, cumulative hazard function, probability density function and the survivor function are shown in figure 3.4.

Table 3.17 Estimated Survival Probabilities for various Distributions

<table>
<thead>
<tr>
<th>Distribution</th>
<th>1-Month</th>
<th>2-Month</th>
<th>3-Month</th>
<th>4-Month</th>
<th>5-Month</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>0.763859</td>
<td>0.583480</td>
<td>0.445696</td>
<td>0.340449</td>
<td>0.260055</td>
</tr>
<tr>
<td>Weibull</td>
<td>0.935342</td>
<td>0.755115</td>
<td>0.521795</td>
<td>0.307182</td>
<td>0.153548</td>
</tr>
<tr>
<td>Lognormal</td>
<td>0.957793</td>
<td>0.723599</td>
<td>0.472635</td>
<td>0.295116</td>
<td>0.183278</td>
</tr>
<tr>
<td>Log logistic</td>
<td>0.951397</td>
<td>0.742228</td>
<td>0.484105</td>
<td>0.297530</td>
<td>0.186014</td>
</tr>
</tbody>
</table>

From the above table observed that, Survival probabilities of different months for lognormal, log-logistic and Weibull distributions are almost similar and seem to be good when compared with exponential distribution. Also as age increases the survival probability is decreases.
Table 3.18 Goodness-of-Fit

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Anderson-Darling (adjusted)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weibull</td>
<td>166.861</td>
</tr>
<tr>
<td>Lognormal</td>
<td>167.657</td>
</tr>
<tr>
<td>Exponential</td>
<td>182.206</td>
</tr>
<tr>
<td>Log logistic</td>
<td>167.67</td>
</tr>
</tbody>
</table>

From the above table observed that, the deviance for lognormal, log logistic and Weibull distributions are almost similar and small when compared with exponential distribution. However the deviance of Weibull distribution, is little less compared to the others, it gives better fit to the study data.

Table 3.19 Mean Survival Times for various Distributions

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Mean</th>
<th>Standard Error</th>
<th>95% Normal CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Lower</td>
</tr>
<tr>
<td>Weibull</td>
<td>3.27218</td>
<td>0.093176</td>
<td>3.09457</td>
</tr>
<tr>
<td>Lognormal</td>
<td>3.47202</td>
<td>0.134688</td>
<td>3.21782</td>
</tr>
<tr>
<td>Exponential</td>
<td>3.71478</td>
<td>0.217764</td>
<td>3.31157</td>
</tr>
<tr>
<td>Log logistic</td>
<td>3.67238</td>
<td>0.151075</td>
<td>3.38791</td>
</tr>
</tbody>
</table>

From the above table observed that, the estimated mean survival times are very similar, at 3.71 months when the times are assumed to have an exponential distribution, 3.27 mont when the times are assumed to have a Weibull distribution, 3.46 months when the times are assumed to have a lognormal distribution and 3.67 months when the times are assumed to have a log-logistic distribution. However, the standard error is small when the times are
assumed to have a Weibull model. The mean is therefore estimated more precisely when the time are assumed to have a Weibull distribution.

3.13 APPLICATION TO CLINICAL TRIAL FOR NON PARAMETRIC METHODS

The data consists of 360 tuberculosis patients admitted in randomized controlled clinical trial into different treatments and the duration of the treatment is six months in each group. The event of interest is sputum smear conversion (positive to negative) during the treatment period. The event is coded as 1 and censoring is coded as 0. The following five covariates are considered for analysis. Time and Status are also included for the analysis.

1. Age (in years)
2. Sex (Male-1 and Female-0)
3. Treatment
4. Weight at baseline (in Kg)
5. Pre treatment susceptibility (Present-1 and Absent-0)

Table 3.20 Mean for Survival Time

<table>
<thead>
<tr>
<th>sex</th>
<th>Mean Estimate</th>
<th>Std. Error</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Lower Bound</td>
</tr>
<tr>
<td>F</td>
<td>2.596</td>
<td>0.17</td>
<td>2.262</td>
</tr>
<tr>
<td>M</td>
<td>3.07</td>
<td>0.108</td>
<td>2.858</td>
</tr>
<tr>
<td>Overall</td>
<td>2.945</td>
<td>0.092</td>
<td>2.765</td>
</tr>
</tbody>
</table>
### Table 3.21 Median for Survival Time

<table>
<thead>
<tr>
<th>sex</th>
<th>Mean Estimate</th>
<th>Std. Error</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Lower Bound</td>
</tr>
<tr>
<td>F</td>
<td>2.000</td>
<td>0.152</td>
<td>1.702</td>
</tr>
<tr>
<td>M</td>
<td>3.000</td>
<td>0.144</td>
<td>2.719</td>
</tr>
<tr>
<td>Overall</td>
<td>2.000</td>
<td>0.121</td>
<td>1.763</td>
</tr>
</tbody>
</table>

### Table 3.22 Comparison Tests for various Models

<table>
<thead>
<tr>
<th>Model</th>
<th>Chi-Square</th>
<th>Df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Rank (Mantel-Cox)</td>
<td>5.700</td>
<td>1</td>
<td>0.017</td>
</tr>
<tr>
<td>Breslow (Generalized Wilcoxon)</td>
<td>6.065</td>
<td>1</td>
<td>0.014</td>
</tr>
<tr>
<td>Tarone-Ware</td>
<td>6.141</td>
<td>1</td>
<td>0.013</td>
</tr>
</tbody>
</table>

### Table 3.23 Life Table

<table>
<thead>
<tr>
<th>Interval</th>
<th>Beg. Total</th>
<th>Convrtd</th>
<th>Cens</th>
<th>Survival</th>
<th>S.E</th>
<th>95% Conf. Int.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-2</td>
<td>360</td>
<td>104</td>
<td>0</td>
<td>0.7111</td>
<td>0.0239</td>
<td>0.6613-0.7550</td>
</tr>
<tr>
<td>2-3</td>
<td>256</td>
<td>167</td>
<td>1</td>
<td>0.2463</td>
<td>0.0227</td>
<td>0.2031-0.2919</td>
</tr>
<tr>
<td>3-4</td>
<td>88</td>
<td>44</td>
<td>4</td>
<td>0.1203</td>
<td>0.0173</td>
<td>0.0890-0.1566</td>
</tr>
<tr>
<td>4-5</td>
<td>40</td>
<td>13</td>
<td>16</td>
<td>0.0714</td>
<td>0.0147</td>
<td>0.0463-0.1037</td>
</tr>
<tr>
<td>5-6</td>
<td>11</td>
<td>2</td>
<td>0</td>
<td>0.0584</td>
<td>0.0146</td>
<td>0.0343-0.0916</td>
</tr>
<tr>
<td>6-7</td>
<td>9</td>
<td>1</td>
<td>8</td>
<td>0.0468</td>
<td>0.0157</td>
<td>0.0225-0.0844</td>
</tr>
</tbody>
</table>

### Table 3.24 Case Processing Summary

<table>
<thead>
<tr>
<th>Sex</th>
<th>Total No</th>
<th>No of Events</th>
<th>Censored</th>
<th>Censored N</th>
<th>Censored Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>94</td>
<td>82</td>
<td>12</td>
<td>12</td>
<td>12.8%</td>
</tr>
<tr>
<td>M</td>
<td>266</td>
<td>214</td>
<td>52</td>
<td>52</td>
<td>19.5%</td>
</tr>
<tr>
<td>Overall</td>
<td>360</td>
<td>296</td>
<td>64</td>
<td>64</td>
<td>17.8%</td>
</tr>
</tbody>
</table>
Figure 3.5 Representation of Life Table

Figure 3.6 Kaplan-Meier Survival Estimates
Survivor functions, by sex adjusted for age

Figure 3.7 Kaplan-Meier Curves for Adjusted Age

Figure 3.8 Kaplan-Meier for Treatments
3.14 DISCUSSION

The estimated median survival time in exponential distribution is 2.885 months and an estimate of 75th percentile of the distribution of survival time is \( \hat{\tau}(75) = 5.1464 \). That is, 75% of Skin TB patient who are undergoing for regular treatment will have chances to get free from a disease of time less than 5 months.

The estimated median survival time in Weibull distribution is 3.09345 months and an estimate of 75th percentile of the distribution of survival time is \( \hat{\tau}(75) = 4.32304 \). That is, 75% of Skin TB patient who were undergone for regular treatment will have chances to get free from the disease of time less than 4.3 months.

The estimated mean survival time in lognormal distribution is \( \bar{\tau} = 3.46955 \) and the standard error of \( \bar{\tau} \) is 0.134068. The estimated
median survival time is 2.87652 months and an estimate of 75th percentile of the distribution of survival time is $\mathbb{E}(75) = 4.34732$. That is, 75% of Skin TB patient who were undergone for regular treatment will have chances to get free from the disease of time less than 4.34 months.

The estimated median survival time in log-logistic distribution is 2.93179 months and an estimate of 75th percentile of the distribution of survival time is $\mathbb{E}(75) = 4.36194$. That is, 75% of Skin TB patient who were undergone regular treatment will have chances to get free from the disease of time less than 4.36 months.

The non-parametric tests including Kaplan-Meier test were performed for this clinical trial data and the outcome shows that there was some difference in survival time mean of estimate male and female. Logrank, Tarone-Ware and Breslow tests show that there is significant difference between survival functions. The Logrank test is used for testing equality of survival functions by weighting all time points which are same. Where as the Tarone-Ware is test equality of survival functions by weighting all time points by the square root of the number of cases at risk at each time point. This is considered a compromise between the log rank and Breslow methods. If we want to compare factors which supports various statistics we need to test for equality of survival function between different treatments. The findings of log rank, Breslow and Tarone-Ware tests show significant differences which means that the survival function is equal between treatments.
Except the treatment 2 all the other graphical representations for the different treatments of observed and predicted values converge in fourth month whereas the graphical representation of observed and predicted values for male and female converge in the sixth month.