CHAPTER II

PARAMETRIC AND SEMIPARAMETRIC REGRESSION MODELS

2.1 Introduction

The early efforts in development of survival analysis were predominantly focused on estimation of the hazard function and the survival function (Cox, 1972). Tsiatis (1975) and Peterson (1976) established that neither the hazard function nor the survival function is identifiable if there are censored observations. Chiang (1968) referred to the hazard function estimated with censored observations as crude hazard and denoted it as \( h(t) \). The actual hazard was called the net hazard. Therefore, in most survival analysis applications, a key assumption is made regarding the equality of the crude hazard (that is estimable) and the net hazard (that is of interest). Various non-parametric survival models and their application have been investigated and recommended in the last two decades.

Semi parametric regression models contain components of parametric and nonparametric components, and the advantages of parametric and nonparametric regression models. In case of an optimal combination of the two approaches, the semi parametric regression model has the efficiency (low variance) of parametric models and the flexibility (small bias) of nonparametric models. However, semi parametric models are not as flexible as non parametric model.

For the past four decades the Cox proportional (PH) hazards model has been used comprehensively to examine the covariate effects on the hazard function for the failure time variable. On the other hand, the accelerated failure time model (AFT), which simply regresses the logarithm of the survival time over the covariates, has been
utilized in the analysis of censored survival data recently. The AFT model has an intuitive physical interpretation and very useful alternative to the Cox model in survival analysis.

2.2 Non-Parametric Methods

Kaplan and Meier’s product limit method is the most commonly used technique for estimating the survivorship function for samples of small, moderate 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 (Venkatesan and Sekar, 2013).

2.2.1 Life Table Method

The simplest method of analyzing survival data is the life table method. Life tables are particularly suited for analyzing very large data sets (Young et al., 1999). 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. 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 cohort life tables and current life tables.

The life table method assumes that subjects are withdrawn randomly throughout each interval therefore; on average they are withdrawn half way through 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 the same for all subjects and is independent of the probability of survival at other time periods. Life 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.

For estimating survival function \( S(t_k) \), the method of construction of life-table is: partitioning the period of observation into a fixed sequence of intervals \( I_1, I_2, ..., I_k \) not necessarily of equal length, but for humans the interval is usually one year. Taking \( n_i, d_i, l_i, w_i \) respectively are the number of individuals alive at the beginning of \( I_i \), died during \( I_i \), lost to follow up during \( I_i \), with drawn during \( I_i \) and

\[
p_i = P\{\text{surviving through } I_i \text{ or alive at the beginning of } I_i \}
\]

and \( q_i = (1 - p_i) \)

The survival probability of \( S(t_k) \) can be written as a product of probabilities

\[
S(t_k) = P(T > t_k) = P(T > t_1) P(T > t_2 | T > t_1) ... P(T > t_k | T > t_{k-1})
\]

\[
= p_1 p_2 ... p_k \tag{2.1}
\]

By reduced sample method we ignore the information that is contained in \( l_i \) and \( w_i \). It is a biased estimate of \( S(t) \).
In Actuarial method, \(1 - \frac{d_i}{n_i}\) is used to estimate \(p_i\) if there are no losses or withdrawals in \(I_i\), 

\[
\hat{q}_i' = \hat{p}_i' = 1 - \hat{q}_i' = \frac{d_i}{n_i}
\]

Actuarial estimate is

\[
\hat{S}(t_k) = \prod_{i=1}^{k} P_i
\]

and

\[
\text{Vár } \hat{S}(t_k) = \hat{S}(t_k)^2 \sum_{i=1}^{k} \frac{d_i}{n_i' (n_i' - d_i)}
\]

This formula is a large sample approximation and it is not reliable for smaller sample size. This estimator was called as the standard life-table estimator and it was initially proposed and explored by Berkson and George (1950) and Cutler and Ederer (1958).

### 2.2.2 Kaplan-Meier Estimator

The Kaplan and Meier or product limit estimator provide a non-parametric maximum likelihood estimate of the survivor function (Kaplan and Meier, 1958). The Kaplan-Meier product limit method is a special case of the lifetable technique, in that series of time intervals is formed in 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 \(t\). The estimate is product of a series of estimated conditional probabilities. The Kaplan Meier estimate of \(S(t)\) is given as

\[
S(t) = \prod_{i=1}^{t} \left[ \frac{(n-i)}{n-i+1} \right]^{C_i}
\]

\(S(t)\) is estimated survival function at time \(t\).

\(\prod_{i=1}^{t} \) denotes the multiplication of the survival times across all cases less 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. where $C_i$ is a constant such that, the code 1 is uncensored case, or terminal case and 0 is coded for censored cases. It can be derived in the form as

$$\hat{S}(t) = \prod_{t_i < t} \frac{n_i - d_i}{n_i} \quad (2.5)$$

where $n_i$ corresponds to the number of observations at risk of failing just prior to time $t_i$ and $d_i$ denotes the number of failures at time $t_i$.

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 zero and is undefined beyond the largest observation, unless an additional assumption is imposed. In adding together, if less than 50% of the observations are uncensored and the largest observation is censored, the median survival time cannot be estimated. Thus the method is not perfect and there are reasons to search for a parametric model.

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

$$\hat{S}(k) = p_1 \times p_2 \times p_3 \times \ldots \times p_k \quad (2.6)$$

37
where $p_1$ denotes the proportion of patients surviving at least one year, $p_2$ denotes the proportion of patients surviving the second year after they have survived one year, $p_3$ denotes the proportion of patients surviving the third year after they have survived two years, and $p_k$ denotes the proportion of patients surviving 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,

$$\hat{S}(t) = \hat{S}(t-1)p_t$$

(2.7)

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

$$\hat{S}(t) = \prod_{t_{(r)}} \frac{n-r}{n-r+1}$$

(2.8)

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 $50^{th}$ percentile, which is the value of $t$ as $\hat{S}(t) = 0.50$.

### 2.3 Tests for Comparing the Survival Curves

#### 2.3.1 Log Rank Test

The log rank test is a large sample chi-square test that uses as 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. While 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, 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. To estimate the log survival function Altshuler (1970) at \( t_{(i)} \) using

\[
-e(t_{(i)}) = -\sum_{j \neq t_{(i)}} \frac{m_{(j)}}{r_{(j)}}
\]

(2.9)

where \( m_{(i)} \) is number of failure and \( r_{(i)} \) number of observations, the score \( w_i = 1 - e(t_{(i)}) \) for uncensored observation \( t_{(i)} \) and \( -e(T) \) for an observation censored at \( T \). The log rank test is based on the sum \( S \) of the \( w \) scores and the permutation variance of \( S \) is

\[
\text{Var}(S) = \frac{n_1 n_2 \sum_{i=1}^{n_1+n_2} w_i^2}{(n_1 + n_2)(n_1 + n_2 - 1)}
\]

(2.10)

This can be rewritten as

\[
V = \left\{ \sum_{j=1}^{k} \frac{m_{(j)}(r_{(j)} - m_{(j)})}{r_{(j)}} \right\} \left( \frac{n_1 n_2}{n_1 + n_2(n_1 + n_2 - 1)} \right)
\]

(2.11)

the statistic \( L = S/\sqrt{\text{Var}(S)} \) has an asymptotically standard normal distribution under the null hypothesis. The log rank statistic \( S \) 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. The test statistic

$$X^2 = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2}$$

(2.12)

### 2.3.2 Peto and Peto’s Generalized Wilcoxon Method

It is a two sample rank sum test and described by (Peto and Peto, 1972). This test seems similar to logrank test, this test assigns a score to every observation. The score is $u_i$ for an uncensored observation $t$, where

$$u_i = \hat{S}(t^+) + \hat{S}(t^-) - 1$$

(2.13)

and for an observed censored at $T$, the score is

$$u_i = \hat{S}(T) - 1$$

(2.14)

These generalized Wilcoxon scores sum to zero. The test procedure after the scores are assigned is the same as for the logrank test.

### 2.3.3 Breslow Test

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 a lognormal survival distribution, but may have low power if there is heavy censoring. The Gehan-Breslow test gives more weight to earlier failures (deaths).

### 2.3.4 Harrington-Flemming Test

The weight used here is $(S(t_{i-1}))^\rho$, where $S(t_{i-1})$ is the Kaplan-Meier estimator of the survival function.
\[
U_i = \sum_{j=1}^{L} (S(t_{i-j})^j (d_{ij} - e_{ij})
\]

(2.15)

Like the Peto and Prentice test, the advantage here is that the weight is related to the overall survival experience. To obtain the assessment of the dependence on the covariates the log-rank method is useful and it is used in for the analysis.

### 2.3.5 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

\[
T = \frac{\sum_{i=1}^{k} w_i (d_{ii} - E_{ii})}{\sum_{i=1}^{k} \left( w_i^2 v_{ii} \right)^{1/2}}
\]

(2.16)

This statistic was may also be 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 T and comparing 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 which is nothing but Mantel-Haenszel Test, Breslow is also called Generalized Wilcoxon test and third one is Tarone-Ware. The equation

\[
U = \sum W_i (D_i - E_i)
\]

(2.17)

where \( W_i \) denotes Weight, \( D_i \) is denoted as number of terminal events observed and \( E_i \) is the number of terminal events expected which is calculated by the number at risk cases & 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 weighted equally and this is considered as the least conservative among the three tests. Where as Breslow test consider as the most conservative of the other tests and then the earlier event weighted more heavily than any other tests also \(W_i\) 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 less heavily than Breslow test does. The \(W_i\) is 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 most powerful than the Breslow test only if the number of terminated cases of the groups is proportional i.e., the mortality of the groups by a constant multiplier. As compared to Breslow is also more powerful than the log-rank test if the mortality of the groups is not proportional. Another important point of the power of the Breslow test is going to declines as the number of censored cases increases.

2.3.6 Partial Likelihood Ratio Test

The partial likelihood ratio test, denoted by \(G\), is calculated as twice the difference between the log partial likelihood of the model containing the covariate and the log partial likelihood of the model not containing the covariate. Specifically,

\[
G = -2(L_p(b) - L_p(0)) \tag{2.18}
\]

where \(b\) is the estimator of \(\beta\) and \(L_p(0) = -\sum \log (n_i)\) for \(i = 1\) to \(k\); \(n_i\) denotes the number of subjects in the risk set at observed survival time \(t_i\). Under the null hypothesis, i.e., that the coefficient is equal to zero, (along with other mathematical conditions), this statistic will follow a chi-square distribution with 1 degree of freedom.
This distribution can be used to obtain p-values to test the significance of the coefficient.

### 2.3.7 Wald Test

This is another test to test for significance of the coefficient which can be computed from the ratio of the estimated coefficient to its estimated standard error. This ratio is commonly referred to as a Wald statistic. The Wald statistic will follow a standard normal distribution and is given by

\[
Z = \frac{b}{\text{SE}(b)}
\]

where \( b \) is the estimator of \( \beta \)

The estimator of the standard error of \( b \) is given as \( \left[I(b)^{-1}\right]^{1/2} \) where \( I(b)^{-1} \) is the observed information matrix. The Wald test is used in our analysis to test the significance of the covariates.

### 2.3.8 Score Test

The test statistic of the score test is the ratio of the derivative of the log partial likelihood equation, to the square root of the observed information all evaluated at \( \beta = 0 \).

\[
Z^* = \frac{\partial \log L}{\partial \beta} \left[ I(\beta)^{-1} \right]^{1/2}, \text{ evaluated at } \beta = 0
\]

Under the hypothesis that the coefficient is equal to zero and the same mathematical conditions required for the Wald and partial likelihood ratio tests, this statistic follows a standard normal distribution.

In practice, the numeric values of the three tests (\( G^2, Z \) and \( Z^* \)) should lead one to draw the same conclusion about the significance of the coefficient. In situations
where there is disagreement, making it necessary to choose one test, the partial likelihood ratio test is the preferred choice.

### 2.4 Proportional Hazard (PH) Models

Semi parametric regression models contain components of parametric and nonparametric regression models, and in some sense combine the advantages and disadvantages of parametric and nonparametric regression models. In case of an optimal combination of the two approaches, the semi parametric regression model has the efficiency (low variance) of parametric models and the flexibility (small bias) of nonparametric models. However, semi parametric models are not as flexible as non parametric model.

Univariate survival data are analyzed by non-parametric methods. The effects of more than one variable require modeling technique to be employed in analysis. In practical situation, the survival of a patient, the survival of a patient depends on the prognostic or explanatory variables. The relationship between the survival experience of a patient and explanatory variables is unified by multifactorial modeling approaches.

#### 2.4.1 Cox proportional hazard (PH) model

In many applications there is a need, when comparing two treatments, to make adjustments for other covariates that may affect outcome. If the two treatments are found to have different survival rates then to attempt comparing the estimated survival functions for the two treatments using a stratified Cox (1972) proportional hazards model.

Cox PH model is an example of a semi-parametric model. The hazard ratio of any two individuals is constant over time and it does not depend on the baseline hazard. Though the Cox PH model is introduced initially in the framework of proportional
hazards functions, Cox PH models also can accommodate covariates that are time dependent and multiple events.

The Cox PH model has been widely used for independently identically distributed right-censored time to event data since 1972. The Cox regression model is a well-recognized statistical technique for exploring the relationship between the survival of a patient and several explanatory variables. This model is referred to in the literature by a variety of terms, such as the Cox model, the Cox proportional hazards model or simply the proportional hazards model. A Cox model provides an estimate of the treatment effect on survival after adjustment for other explanatory variables. It allows us to estimate the hazard (or risk) of death, or other event of interest, for individuals, given their prognostic variables. Even if the treatment groups are similar with respect to the variables known to effect survival, using the Cox model with these prognostic variables may produce a more precise estimate of the treatment effect (for example, by narrowing the confidence interval). Interpreting a Cox model involves examining the coefficients for each explanatory variable. A positive regression coefficient for an explanatory variable means that the hazard is higher, and thus the prognosis worse, for higher values. Conversely, a negative regression coefficient implies a better prognosis for patients with higher values of that variable.

The test for the non-violation of PH assumption can either be done formally or graphically. The formal test is done by interacting each covariate of interest with the log (time) variable. Using Log (time) instead of time variable ensures that there is no numerical overflow. A non-significant coefficient in the interaction covariate means that the PH assumption is satisfied. A non-significant P-value for the correlation coefficients means there is no trend.
The model where the hazard function of death at a particular time depends on the values $X_1, X_2, \ldots, X_p$ of $p$ explanatory variables $X_1, X_2, \ldots, X_p$. The values will be assumed to record at the time of study. The set of values of the explanatory variables in the proportional hazard model will be represented by the vector $X$, so that $X = (X_1, X_2, \ldots, X_p)'$. Let $h_0(t)$ be the hazard function for an individual for whom the values of all the explanatory variables that make up the vector $X$ are zero. The function $h_0(t)$ is called the base line hazard function. The hazard function for the $i^{th}$ individual can be written as

$$h_i(t) = h_i(t)h_0(t)$$

where

$$h_i(t) = \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$$

The general class of PH models is

$$h(t, z) = h_0(t)g(z\beta)$$

where $h_0(t)$ is the base line hazard at time $t$ and $z$ is the covariate. Cox (1972) specified $g(z\beta) = e^{z\beta}$ which is independent of $t$ giving the ratio of the hazard constant over time.

The time dependent Cox model is

$$h(t, z(t)) = h_0(t)e^{z(t)\beta}$$

this is a no longer PH model.

Cox regression aims to estimate the hazard ratio and it gains power and precision and also it can accommodate both discrete and continuous measures of event times. Cox models the effect of covariates on the hazard rate but leaves the baseline hazard rate unspecified and it does not assume knowledge of absolute risk. Also it estimates relative rather than absolute risk. The proportional hazards model Cox (1972) specifies that the cumulative hazard function for the survival time associated with possibly time-dependent covariates $Z$ takes the form
\[
\Delta(t \mid Z) = \int_0^t \exp(\beta_0^T Z(s)) \, d\Delta_0(s)
\]  
(2.24)

where \( \beta_0 \) is a vector-valued regression parameter, and \( \Lambda_0(\cdot) \) is an unspecified baseline cumulative hazard function. Suppose that we have a random sample of \( n \) subjects. For \( i = 1, \ldots, n \) let \( T_i \) be the survival time, \( C_i \) be the censoring time, and \( Z_i(\cdot) \) be the vector of covariates. \( \overline{T}_i = \min(T_i, C_i) \), \( \delta_i = I(T_i \leq C_i) \), \( N_i(t) = \delta_i I(\overline{T}_i \leq t) \) and \( Y_i(t) = I(\overline{T}_i \geq t) \)

where \( I(\cdot) \) is the indicator function. Suppose that \( C \) is independent of \( T \) conditional on \( Z \) and that the data are observed on the time interval \((0, \tau)\), where \( 0 < \tau < 1 \). Then a consistent estimator \( \hat{\beta} \) of \( \beta_0 \) can be obtained by solving the partial likelihood score equation \( U(\beta) = 0 \), where

\[
U(\beta) = \sum_{i=1}^n \int_0^t \left\{ Z_i(t) - \frac{\sum_{j=1}^n Y_j(t) \exp(\beta^T Z_j(t)) Z_j(t)}{\sum_{j=1}^n Y_j(t) \exp(\beta^T Z_j(t))} \right\} \, dN_i(t) 
\]  
(2.25)

The baseline cumulative hazard \( \Delta_0(t) \) can be estimated by

\[
H_0(t) = \sum_{i=1}^n \int_0^t \frac{dN_i(s)}{\sum_{j=1}^n Y_j(s) \exp(\hat{\beta}^T Z_j(t))} 
\]  
(2.26)

Notice that there is no constant term in linear component of the proportional hazard model. If a constant \( \beta_0 \) is included, the base line hazard function could simply be recalled by dividing \( h_0(t) \) by \( \exp(\beta_0) \), and the constant term would cancel out. Moreover, we have made no assumptions concerning the actual form of the base line hazard function \( h_0(t) \). Indeed, we see that the \( \beta \) coefficients in this proportional hazard model can be estimated without making any such assumptions. Once \( \beta \) has been estimated through maximum likelihood, then the hazard function for an individual with covariate \( x \) is estimated via
\[ \hat{h}_x(y) = \hat{h}_0(y)e^{\hat{\beta}'\xi} \]  

(2.27)

where \( \hat{\beta}' \) is the Cox regression estimator of \( \beta \). This requires that the baseline hazard \( h_0(y) \) must first be estimated. Equivalently, the baseline survival function \( s_0(y) \) may be estimated by \( \hat{S}_0(y) \) using maximum likelihood methods as outlined by Kalbfleisch and Prentice (1973). Finally,

\[ \hat{S}_0(y) = \prod_{j=1}^{u} \hat{\xi}_j, \text{ for } y_{(u)} \leq y < y_{(u+1)}, \quad u=1,2,\ldots,k-1 \]  

(2.28)

The basic assumption of Cox proportional hazard model is the hazard ratio is constant across time and the observations are independent.

### 2.4.2 Extended Cox PH Models

Time varying covariates are usually classified as being either internal or external. An internal time varying covariate is one whose value is subject-specific and requires that the subject be under periodic observation. An external time varying covariate is one whose value at a particular time does not require subjects to be under direct observation.

Let \( X(t) \) denote the value of the covariate \( x \) measured at time \( t \). Let \( x_l(t_i) \) denote the value of the covariate for subject \( l \) at time \( t_i \). To allow for multiple covariates, we let \( x_{lk}(t_i), k=1,2,\ldots,p \) denote the value of the \( k^{th} \) covariate for subject \( l \) at time \( t_i \) and denote the vector of covariates as

\[ x_l(t_i) = [x_{l1}(t_i), x_{l2}(t_i), \ldots, x_{lp}(t_i)] \]  

(2.29)

The notation in above equation is completely general in the sense that, if a particular covariate, \( x_k \), is fixed then

\[ x_{lk}(t_i) = x_{lk}(t=0) = x_{lk} \]  

(2.30)

The generalization of the proportional hazards regression function
to include possibly multiple time varying covariates is

\[
h(t, x(t), \beta) = h_0(t) \exp[x'\beta(t)]
\]  \hspace{1cm} (2.32)

and the generalization of the partial likelihood function

\[
I_p(\beta) = \prod_{i=1}^{n} \left[ \frac{e^{x_i'\beta}}{\sum_{j \in R_i} e^{x_j'\beta}} \right]
\]  \hspace{1cm} (2.33)

The Cox Proportional-hazards regression model has achieved extensive use in the analysis of time-to-event data with censoring and covariates. Time varying covariates are especially tractable within the Cox model and are little more complex than that for time-fixed covariates. Indeed for covariates that change rapidly over time the Cox model is often simpler to manage than a parametric survival model.

Let \( x_i(t) \) denote the value of a vector of covariates for individual at time \( t \). Then the proportional hazards model may be generalized to

\[
h_i(t, x_i(t)) = h_0(t) \exp\{x_i'\beta(t)\}
\]  \hspace{1cm} (2.34)

This model allows for great generality. In the two-sample case, for example, the model may be written as

\[
h_i(t \mid x) = \begin{cases} h_0(t) & \text{if } x = 0 \\ h_0(t)e^{\beta(t)} & \text{if } x = 1 \end{cases}
\]  \hspace{1cm} (2.35)

This basically allows for two arbitrary hazard functions, one for each group. Thus, this is a form of saturated model. Usually the form of time dependence of the effects must be specified parametrically in order to be able to identify the model and estimate the parameters. Obvious candidates are polynomials on duration, where \( \beta(t) \) is
a linear or quadratic function of time. (Cox and Oakes, 1984) show how one can use quick-dampening exponentials to model transient effects.

It is tempting therefore to consider making repeated measures of them during the operation of the trial and by fitting a model that uses these updated covariate values. The survival period of each patient to be divided up into a sequence of shorter survival spells, each characterized by an entry time and an exit time, and within which covariate values remain fixed.

2.5 Accelerated Failure Time (AFT) Regression Model

The AFT model is an attractive alternative to the popular Cox proportional regression model. The accelerated failure time (AFT) model is a linear regression model in which the response variable is the logarithm or a known monotone transformation of a failure time (Kalbfleisch and Prentice, 1980). There are literatures for the different types and several approaches of AFT have been proposed for the estimation and inference on the model. Semi parametric estimation in the AFT model with an unspecified error distribution has been studied extensively in the literature for right censored data. In particular, two methods have received special attention. One method is the Buckley-James estimator which adjusts censored observations using the Kaplan-Meier estimator. The other is the rank based estimator which can be motivated from the score function of the partial likelihood, described by many authors (Prentice, 1978; Buckley and James, 1979; Ritov, 1990; Tsiatis, 1990; Wei et al., 1990; Ying, 1993). Rank-based methods were studied (Tsiatis, 1990; Wei and Tanner, 1990; Lai and Ying, 1991; Lai and Ying, 1992; Lin and Geyer, 1992; Ying, 1993; Fygenson and Ritov, 1994). Least squares based and M-estimation methods were investigated by (Miller, 1976; Buckley and James, 1979; Koul et al., 1981; Ritov, 1990; Lai and Ying, 1991). The inference methods for the AFT model for right censored data were
described by Jin et al., (2003). The AFT model is a linear regression model in which the response variable is the logarithm or a known monotone transformation of a failure time (Kalbfleisch and Prentice, 1980).

Let $T_i$ be a random variable denoting the failure time for the $i^{th}$ subject, and let $x_{i1}, x_{i2}, \ldots, x_{ip}$ be the values of $p$ covariates for that same subject. The model is

$$\log T_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} + \sigma \epsilon_i ; \epsilon_i \sim S_0(.)$$

(2.36)

where $\epsilon_i$ is a random disturbance term, and $\beta_0, \ldots, \beta_p$, and $\sigma$ are parameters to be estimated, $S_0(\cdot)$ is a known baseline survival, $T_i$ is actual survival time sometimes observed, $\sigma$ is a scale parameter and $x_i$'s are fixed $p \times 1$ vector of covariates. The $\sigma$ can be omitted, which requires that the variance of $\epsilon_i$ be allowed to be different from one. But it is simple to fix the variance of $\epsilon_i$ at 1 and let $\sigma$ change. All AFT models are named for the distribution of $T$ rather than the distribution of $\epsilon$ or $\log T$. The reason for allowing different distribution assumptions is that they have different implications for the shapes of hazard function.

### 2.6 Parametric Regression Models

Parametric regression models play an important role in survival analysis, since many analyses in practice are carried out using parametric models. Parametric modeling offers straightforward modeling and analysis techniques. Usually, there are many physical causes that lead to the failure or death of an individual. It is very difficult, if not impossible, to isolate these physical causes and mathematically account for all of them. Therefore, choosing a theoretical distribution to approximate survival data is as much an art as a scientific task. In this chapter, several theoretical distributions that have been widely used to describe survival time are discussed, their characteristics summarized, and their applications illustrated. The parametric models
have some advantages. In particular (a) full maximum likelihood may be used to estimate the parameters, (b) the coefficients can be clinically meaningful and, for some models, are related to those from a proportional model, (c) fitted values from the model can provide estimates of survival time and (d) residuals can be computed that are differences between observed and predicted values of time. The result is that an analysis using a fully parametric model can have the look and feel of a normal errors linear regression analysis.

There are a number of discussions about parametric regression model by different authors in different fields (Crowder et al., 1991; Cox and Oakes, 1984; Elandt-Johnson and Johnson, 1980; Gross and Clark, 1975; Lawless, 1982; Lee, 1992a; Nelson, 1982, 1990). (Collett, 1994) suggested that parametric models for the level of comparing at slightly to the higher level of mathematics. Whereas (Andersen et al., 1993) discussed theoretical parts of parametric regression models.

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 contrast 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, is with the equation

$$ t = e^{(b_0 + b_1x)}e^\sigma $$

(2.37)

We express survival time, t, as the product of the systematic component of the model, $e^{(b_0 + b_1x)}$, and the error component $\varepsilon$. This model can be “linearized” by taking the natural log of each side of the equation, yielding
\[ \ln(t) = \beta_0 + \beta_1 X + \varepsilon^* \]  

(2.38)

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

### 2.6.1 Exponential Regression 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 1940’s. 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 Davis (1952). (Epstein and Sobel, 1953) reported the reasons for selecting the exponential distribution over the popular 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 own 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.

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

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

(2.39)

the cumulative distribution function is

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

(2.40)
and the survivorship function and hazard function respectively

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

Hazard function is a constant and independent of t. We consider the single covariate model where 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

\[ S(t, x, \beta) = \exp(-t/e^{\beta_1 x}) \]  

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

\[ t_{50}(x, \beta) = -e^{\beta_0 + \beta_1 x} \times \ln(0.5) \]  

when the covariate is dichotomous, coded 0 or 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

\[ (x = 1, x = 0) = \frac{t_{50}(x = 1, \beta)}{t_{50}(x = 0, \beta)} = \frac{-e^{\beta_0 + \beta_1 x} \times \ln(0.5)}{-e^{\beta_0} \times \ln(0.5)} = e^{\beta_1} \]  

Alternatively, the relationship between the two median times is

\[ t_{50}(x = 1, \beta) = e^{\beta_1} t_{50}(x = 0, \beta) \]  

An alternative way to present multiplicative effect is via the survivorship function. The relationship between the survivorship function for the two groups is

\[ S(t, x = 1, \beta) = S(te^{-\beta_1}, x = 0, \beta) \]  

The interpretation of the result for the above equation 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 \(t\). The change in the sign of the coefficient is due to the fact that time percentiles and survival probabilities are inverse operations of one another.

The hazard function for the model, as shown in

\[
h(t, x, \beta) = e^{-(\beta_0 + \beta x)}
\]  

(2.48)

and the hazard ratio for the dichotomous covariate at \((x=1, x=0)\) is \(e^{\beta_0}\).

We assume 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 \((t_i, x_i, c_i), i = 1, 2, \ldots, n\), the log-likelihood function is

\[
L(\beta) = \sum_{i=1}^{n} c_i z_i - e^{z_i}
\]  

(2.49)

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

\[
\frac{\partial^2 L(\beta)}{\partial \beta_j \partial \beta_k} = -\sum_{i=1}^{n} x_{ij} x_{ik} e^{z_i}, \quad j, k = 0, 1, \ldots, p
\]  

(2.50)

Estimators are based on the observed information matrix, denoted \(I(\hat{\beta})\), which is the matrix with elements given by the negative of the above equation evaluated at the estimator of the coefficients. The inverse of the observed information matrix provides the estimators of the variances and covariances, namely

\[
\hat{\text{Var}}(\hat{\beta}) = I(\hat{\beta})^{-1}
\]  

(2.51)
The end points of a 100(1-\(\alpha\)) percent Wald-statistic-based confidence interval for the \(j^{th}\) coefficients are

\[
\hat{\beta}_j \pm z_{1-\alpha/2} \hat{\text{Se}}(\hat{\beta}_j)
\]

(2.52)

where \(\hat{\text{Se}}(\hat{\beta}_j)\) 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 as

\[
h_\text{exp}(t, \hat{\beta}) = \exp\left(-\hat{\beta}_0\right)
\]

(2.53)

2.6.2 Weibull Regression 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

\[
\ln(T) = \beta_0 + \beta_1 x + \varepsilon^*
\]

(2.54)

The main difference between the Weibull and the exponential regression model is that the parameter \(\sigma\) 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 we begin the Weibull model in the single covariate setting to compare and contrast it with the exponential 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
\[ h(t, x, \beta, \lambda) = \frac{\lambda t^{\beta-1}}{(e^{\beta_0 + \beta_1 x})^{\lambda}} \]  

(2.55)

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

\[
\begin{align*}
  h(t, x, \beta, \lambda) &= \lambda t^{\beta-1} e^{-\lambda(\beta_0 + \beta_1 x)} \\
  &= \lambda t^{\beta-1} e^{-\lambda \beta_1 x} e^{-\lambda \beta_1} \\
  h(t, x, \beta, \lambda) &= \lambda e^{\lambda \beta_1 x} \\
  &= h_0(t) e^{\beta_1 x}
\end{align*}
\]

(2.56)

(2.57)

where \( \gamma = \exp(-\beta_0 / \sigma) = \exp(\theta_0) \), \( \theta_1 = -\beta_1 / \sigma \) and the baseline hazard function is \( h_0(t) = \lambda \gamma t^{\beta-1} \). Although the parameter \( \sigma \) is a variance-like parameter on the log-time scale, \( \lambda = 1/\sigma \) is commonly called the shape parameter and \( \gamma \) is called the scale parameter, then it can be re-expressed as

\[
\begin{align*}
  h(t, x, \beta, \lambda) &= \lambda t^{\beta-1} e^{-\lambda(\beta_0 + \beta_1 x)} \\
  &= \lambda e^{\lambda \beta_1 x} (te^{\beta_1 x})^{\lambda-1} e^{-\lambda \beta_1 x}
\end{align*}
\]

(2.58)

The survivorship function corresponding to the accelerated failure time form of the hazard function is

\[
S(t, x, \beta, \sigma) = \exp\left(-t^\lambda \exp\left[-\frac{1}{\sigma}(\beta_0 + \beta_1 x)\right]\right)
\]

(2.59)

we obtained the equation for the median survival time by setting the survivorship function equal to 0.5 and solving for time yielding

\[
t_{50}(x, \beta, \sigma) = \left[-\ln(0.5)\right]^\lambda e^{\beta_0 + \beta_1 x}
\]

(2.60)

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

\[
(x = 1, x = 0) = \frac{t_{50}(x = 1, \beta, \sigma)}{t_{50}(x = 0, \beta, \sigma)}
\]

57
The equation for the log-likelihood function for a simple possibly containing right-censored data is obtained using

\[ z_i = \frac{y_i - x_i \beta}{\sigma} \]  \hspace{1cm} (2.62)

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

\[ L(\beta, \sigma) = \sum_{i=1}^{n} c_i (-\ln(\sigma) + z_i) - e^{z_i} \]  \hspace{1cm} (2.63)

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

\[ \frac{\partial L(\beta, \sigma)}{\partial \beta_j} = \sum_{i=1}^{n} \frac{-x_i}{\sigma} \left( c_i - e^{z_i} \right) = 0, \quad j = 0, 1, 2, \ldots, p \] \hspace{1cm} (2.64)

The score equation for the shape parameter, \( \sigma \), is

\[ \frac{\partial L(\beta, \sigma)}{\partial \sigma} = -\frac{m}{\sigma} \sum_{i=1}^{n} \frac{z_i}{\sigma} \left( c_i - e^{z_i} \right) = 0 \] \hspace{1cm} (2.65)

The individual elements of this matrix to be evaluated are

\[ -\frac{\partial^2 L(\beta, \sigma)}{\partial \beta_j \partial \beta_k} = \frac{1}{\sigma^2} \sum_{i=1}^{n} x_i x_k e^{z_i}, \]

\[ -\frac{\partial^2 L(\beta, \sigma)}{\partial \beta_j \partial \sigma} = \frac{1}{\sigma^2} \sum_{i=1}^{n} x_i z_i e^{z_i}, \]  \hspace{1cm} (2.66)

\[ -\frac{\partial^2 L(\beta, \sigma)}{\partial \sigma \partial \sigma} = \frac{m}{\sigma^2} + \frac{1}{\sigma^2} \sum_{i=1}^{n} z_i^2 e^{z_i}. \]

When evaluated at the solution to the likelihood equations, the information matrix may be expressed as
where $X$ is an $n$ by $p+1$ matrix containing the values of the covariates, $V = \text{diag}(e^{\hat{z}_i})$, an $n$ by $n$ diagonal matrix, and $\hat{z}'=(\hat{z}_1', \hat{z}_2', \ldots, \hat{z}_n')$ with $\hat{z}_i = \frac{y_i - x_i'\hat{\beta}}{\hat{\sigma}}$.

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

$$\hat{\text{Var}}(\hat{\beta}, \hat{\sigma}) = [I(\hat{\beta}, \hat{\sigma})]^{-1}$$

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

$$\hat{M}_i = c_i - \exp(\hat{z}_i)$$

$$= c_i - t_i^\hat{\lambda} \exp(\hat{\lambda} x_i' \hat{\beta})$$

where $\hat{\lambda} = 1/\hat{\sigma}$. The estimator of the cumulative hazard function for the Weibull regression model is

$$\hat{H}(t_i, x_i' \hat{\beta}, \hat{\sigma}) = \exp\left(\frac{y_i - x_i' \hat{\beta}}{\hat{\sigma}}\right)$$

$$= \exp(\hat{z}_i)$$

$$= (t_i, e^{-x_i' \hat{\beta}})$$

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

$$\hat{H}(t_i, x_i' \hat{\beta}, \hat{\sigma}) = c_i - \hat{M}_i$$

The estimator of the baseline hazard for the Weibull model is

$$h_0(t, \hat{\beta}, \hat{\sigma}) = \frac{1}{\hat{\sigma}} \exp(-\beta_0/\hat{\sigma}) t^{(y_i'\hat{\beta}-1)}$$

2.6.3 Log-Logistic Regression 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 the 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,

\[
f(t) = \frac{\alpha \gamma t^{\gamma-1}}{(1 + \alpha t^{\gamma})^2}
\]

(2.73)

\[
S(t) = \frac{1}{1 + \alpha t^{\gamma}}
\]

(2.74)

\[
h(t) = \frac{\alpha \gamma t^{\gamma-1}}{1 + \alpha t^{\gamma}}
\]

(2.75)

\[
H(t) = \log(1 + \alpha t^{\gamma}), \quad t \geq 0, \quad \alpha > 0, \quad \gamma > 0
\]

(2.76)

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

\[
\ln(t) = \beta_0 + \beta_1 x + \sigma \varepsilon
\]

(2.77)

where the error term $\varepsilon$ 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, we begin by expressing the survivorship function for the model as

\[
S(t,x,\beta,\sigma) = [1 + \exp(z)]^{-1}
\]

(2.78)

where $z = (y - \beta_0 - \beta_1 x)^{-1} / \sigma$ and $y = \ln(t)$.

The odds of a survival time of at least $t$ is

60
\[ \frac{S(t,x,\beta,\sigma)}{1-S(t,x,\beta,\sigma)} = \exp(-z) \]  

(2.79)

As an example, assume that the covariate is dichotomous and coded 0 or 1. The odds-ratio at time \( t \) is evaluated at \( x=1 \) and \( x=0 \) is given as

\[
\text{OR} = \frac{\exp\left[\frac{-\left(y - \beta_0 - \beta_1 \times 1\right)}{\sigma}\right]}{\exp\left[\frac{-\left(y - \beta_0 - \beta_1 \times 0\right)}{\sigma}\right]} = \exp\left(\frac{\beta_1}{\sigma}\right)
\]

(2.80)

An alternative interpretation is obtained when we express the median survival time as a function of the regression coefficients. Setting the survivorship function for the above equation equal to 0.5 and solving, obtain an equation for the median survival time of

\[ t_{50}(x,\beta,\sigma) = \exp(\beta_0 + \beta_1 x) \]

(2.81)

and the time ratio at the median is

\[ (t_{50}, x = 1, x = 0) = \exp(\beta_1) \]

(2.82)

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 is

\[ t_p(x,\beta,\sigma) = \left[\left(1 - p\right)/p\right]^\beta \exp(\beta_0 + \beta_1 x) \]

(2.83)

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

\[ \frac{(1/\sigma)\exp(z)}{[1 + \exp(z)]^2} \]

(2.84)

and the censored time is
where, for a multivariable model, \( z = (y - x'\beta)/\sigma \). It follows that the log-likelihood function for a sample of \( n \) independent observations of time, covariates and censoring indicator, denoted \((t_i, x_i, c_i), i = 1,2,...,n\), is

\[
L(\beta, \sigma) = \prod_{i=1}^{n} c_i \left\{ -\ln(\sigma) + z_i - 2\ln[1 + \exp(z_i)] \right\} - (1 - c_i)\ln[1 + \exp(z_i)]
\] (2.86)

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

\[
\hat{H}(t_i, \hat{\beta}, \hat{\sigma}) = \ln[1 + \exp(\hat{z}_i)]
\] (2.87)

where \( \hat{z}_i = (y_i - x_i'\hat{\beta})/\hat{\sigma} \) and \( \hat{\beta} \) and \( \hat{\sigma} \) denote the estimators maximizing the equation.

### 2.6.4 Lognormal Regression 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 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 (Feinleib and MacMohan, 1960) and Feinleib (1960) have observed that the distribution of survival times of several diseases. Horner (1987) showed 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 variance $\sigma^2$. $T$ is lognormally distributed and can be rewritten as $\Lambda(\mu, \sigma^2)$ where $\mu, \sigma^2$ are two parameters of lognormal distribution.

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

The popularity of lognormal distribution is to the fact that the cumulative values of $y = \log e^t$ 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

$$f(t) = \frac{1}{t\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(\log e^t - \mu)^2\right] ; t > 0, \sigma > 0 \quad (2.88)$$

and

$$S(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{t} \frac{1}{x} \exp\left[-\frac{1}{2\sigma^2}(\log e^x - \mu)^2\right] dx \quad (2.89)$$

Let $a = \exp (-\mu)$. Then $-\mu = \log e \ a$, from the above equations can be written as

$$f(t) = \frac{1}{t\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(\log e^t a)^2\right] \quad (2.90)$$

and

$$S(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{t} \frac{1}{x} \exp\left[-\frac{1}{2\sigma^2}(\log e^x a)^2\right] \frac{dx}{x}$$

$$= 1 - G(\frac{\log e^t a}{\sigma}) \quad (2.91)$$

where $G(y)$ is the cumulative distribution function of a standard normal variable

$$G(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-u^2/2} du \quad (2.92)$$
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 is

$$h(t) = \frac{1}{t\sigma\sqrt{2\pi}} \exp\left(-\frac{(\log_e t)^2}{2\sigma^2}\right) \frac{1}{1 - G(\log_e t / \sigma)}$$

(2.93)

The mean and variance of two parameter lognormal distribution are

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

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

2.6.5 Chi Bar

The likelihood ratio test that is displayed is testing on the boundary of the parameter space. Most likely, you are testing whether an estimated variance component (something that is always greater than zero) is different from zero using an LR test.

Suppose for now that the two models being compared differ only with respect to the variance component in question, in which case the test statistic will be displayed as "chibar(01)". In such cases, the limiting distribution of the maximum-likelihood estimate of the parameter in question is a normal distribution that is halved, or chopped-off at the boundary zero in this case. As a result the distribution of the LR test statistic is not the usual chi-square with 1 degree of freedom, but instead a 50:50 mixture of a chi-square with no degrees of freedom (i.e., a point mass at zero) and a chi-square with 1 degree of freedom.

The p-value of the LR test takes this into account, and will be set to 1 if it is determined that your estimate is close enough to zero to be, in effect, zero for purposes
of significance. Otherwise, the p-value displayed is set to one-half of the probability that a chi-square with 1 degree of freedom is greater than the calculated LR test statistic.

Sometimes you are testing whether a variance component is zero in addition to testing whether k other parameters (not affected by boundary conditions) are zero. Such situations often arise when comparing models fit by xtmixed. For such tests, the distribution of the likelihood-ratio test statistic is a 50:50 mixture of chi-square distributions with k and k+1 degrees of freedom, shown on the output as "chibar(4_5)", for example. As was the case with chibar(01), significance levels are adjusted accordingly.

Finally, if you are testing more than one boundary-affected parameter, the theory is much more complex and, in most cases, intractable. When this occurs, Stata will either display significance levels that are conservative and marked as such, or not display an LR test at all.

2.7 Model Selection Criteria

It is likely that we will have data on more covariates than we can reasonably expect to include in the model, it is to decide on a method to select a subset of the total number of covariates, Venkatesan and Sekar (2011). Often, many covariates are collected and to reduce possible modeling bias, a large parametric model is built. An important and challenging task is to efficiently select a subset of significant variables upon which the hazard function depends, Kuk (1984). There are many variable selection techniques in linear regression models. Some of them have been extended to the context of censored survival data analysis, such as the best subset variable selection and stepwise deletion, Collet (1994).
The modeling approach for the survival data is developed for the dependence of the hazard function on one or more explanatory variables. In this development process, proportional hazard models with linear components that contain different sets of terms are fitted, and comparisons made between them. The common hazard function under the model without introducing any explanatory variables is referred to as the null model.

The general proportional hazards model

\[
 h(t, X) = h_0(t)g(X\beta) \tag{2.94}
\]

In general, suppose that two models are contemplated for a particular data set, where model one contains a subset of the terms in model two. The hazard function for the model with \(p\) explanatory variables is

\[
 h(t, X) = h_0(t)\exp(\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p) \tag{2.95}
\]

the hazard function for model with \(p+q\) explanatory variables is

\[
 h(t, x) = h_0(t)\exp(\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \beta_{p+1} x_{p+1} + \ldots + \beta_{p+q} x_{p+q}) \tag{2.96}
\]

The statistical problem is to determine whether the additional \(q\) terms in model (2.96) significantly improve the explanatory power of the model. If not they may be omitted, and model (2.98) would be deemed to be adequate. The model selection process is to identify a set of explanatory variables that have the potential for being included in the linear component of a PH model. Hazard function may not depend on a unique combination of variables; there must be number of equally good models rather than a single best model. Hence it is desirable to construct a wide range of possible models selection.

The proportional hazards (PH) model is a regression model with duration as dependent variable. It involves the statistical strategies for predictions are similar to
those utilized in ordinary regression. In this model, the techniques of Cox proportional hazards model development is about modeling time to event and its relationship to a set of one or more explanatory variables in the presence of censoring. The Hazard ratio is described parametrically and the baseline hazard function non-parametrically. The feature of the model is that the hazard ratio for any two individuals does not depend on time. We explore the influence of explanatory variables on the hazard ratio through estimation of and test of hypotheses about the regression parameters $\beta_1, \beta_2, \ldots, \beta_p$.

### 2.7.1 Deviance (-2LL)

It is a Measure of agreement between the model and the data. For comparing the alternative nested model fitted to the set of survival data, a statistic which measures the extent to which the data are fitted for the particular model selected. Since the likelihood function summarizes the information that the data contain about the unknown parameters in a given model, a suitable summary statistic is the value of the likelihood function when parameters are replaced by their maximum likelihood estimates. It is more convenient to use minus twice the logarithm of the maximized likelihood in comparing alternative models. If the maximized likelihood for a given model is denoted by $L$, the summary measure of agreement between the model and the data is $-2\log L$ and will always be positive, and for a given data set, the smaller the value of $-2\log L$, the better the model.

### 2.7.2 AIC and BIC

An approach advocated as useful in model building, is one due to Akaike (1973) information criterion (AIC) which examines the likelihood and the number of
parameters included in the model. It attempts to balance the need for a model which fits the data very well to that of having a simple model and penalizes a model for having too many parameters. This serves an alternative to significance tests to estimate quantities of interest, which not necessarily is a nested model and Criterion for choosing between competing statistical models. It judges a model by how close its fitted values tend to be to the true values.

The AIC selects the model that minimizes:

$$AIC = -2 \log L + 2k$$  \hspace{1cm} (2.97)

where $k$ is the number of unknown $\beta$ coefficients.

The AIC count for final sample size

$$AIC_c = AIC + \frac{2k(k+1)}{n-k-1}$$  \hspace{1cm} (2.98)

The other criterions used for model comparison are Bayesian Information Criterion (BIC) and it is closed related to AIC. The penalty term in BIC is larger than AIC. The BIC selects the model that minimizes

$$BIC = -2 \log L + k \ln(n)$$  \hspace{1cm} (2.99)

For the model selection procedure, the statistic $-2 \log L$ and AIC are the better tools to identify the subsets of covariates when adjusting more covariates simultaneously. It is an indicator for models comparisons for the clinical trial set up data. It balanced the need for a model which fits the data very well to that of having a simple model and penalizes a model for having too many parameters.

2.8 Application to Clinical Trial Data
The data consists of 336 HIV infected Tuberculosis patients admitted in clinical trial who were treated with two treatments of six months duration which has been described in Chapter I. The event of interest is death during treatment and follow up period. The event of interest is coded as 1 and censoring is coded as 0.

The covariates considered for models are, Age (years), Sex (Male-1 and Female-0), Treatment group (6 months vs 9 months duration), Weight at baseline (Kg) and CD4 counts (<200-0 and ≥200-1).
Table-2.1 Life table for Comparison of Treatments

<table>
<thead>
<tr>
<th>Covariates</th>
<th>Interval Time</th>
<th>Number Entering Interval</th>
<th>Number Withdrawing during Interval</th>
<th>Number Exposed to Risk</th>
<th>Number of Terminal Events</th>
<th>Proportion Surviving</th>
<th>Cumulative Proportion Surviving at End of Interval</th>
<th>Std. Error of Cumulative Proportion Surviving at End of Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regimen 6M</td>
<td>0-12</td>
<td>169</td>
<td>27</td>
<td>155.5</td>
<td>19</td>
<td>0.88</td>
<td>0.88</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>12-24</td>
<td>123</td>
<td>20</td>
<td>113</td>
<td>11</td>
<td>0.90</td>
<td>0.79</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>24-36</td>
<td>92</td>
<td>9</td>
<td>87.5</td>
<td>9</td>
<td>0.90</td>
<td>0.71</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>36-48</td>
<td>74</td>
<td>9</td>
<td>69.5</td>
<td>2</td>
<td>0.97</td>
<td>0.69</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>48-60</td>
<td>63</td>
<td>5</td>
<td>60.5</td>
<td>3</td>
<td>0.95</td>
<td>0.66</td>
<td>0.04</td>
</tr>
<tr>
<td>Regimen 9M</td>
<td>0-12</td>
<td>167</td>
<td>28</td>
<td>153</td>
<td>19</td>
<td>0.88</td>
<td>0.88</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>12-24</td>
<td>120</td>
<td>14</td>
<td>113</td>
<td>17</td>
<td>0.85</td>
<td>0.74</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>24-36</td>
<td>89</td>
<td>7</td>
<td>85.5</td>
<td>6</td>
<td>0.93</td>
<td>0.69</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>36-48</td>
<td>76</td>
<td>10</td>
<td>71</td>
<td>4</td>
<td>0.94</td>
<td>0.65</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>48-60</td>
<td>62</td>
<td>3</td>
<td>60.5</td>
<td>2</td>
<td>0.97</td>
<td>0.63</td>
<td>0.04</td>
</tr>
</tbody>
</table>

The Table-2.1 shows that the survival probabilities of the years of 1,2,3,4 and 5 for the regimen 6M were 0.88, 0.79, 0.71, 0.69, 0.66 and regimen 9M were 0.88, 0.74, 0.69, 0.65, 0.63 respectively. It shows that the similar survival probabilities for the regimen 9M and regimen 6M.
Table-2.2 Life table for Comparison of CD4 Counts

<table>
<thead>
<tr>
<th>Covariates</th>
<th>Interval</th>
<th>Start Time</th>
<th>Number Entering Interval</th>
<th>Number Withdrawing during Interval</th>
<th>Number Exposed to Risk</th>
<th>Number of Terminal Events</th>
<th>Proportion Surviving</th>
<th>Cumulative Proportion Surviving at End of Interval</th>
<th>Std. Error of Cumulative Proportion Surviving at End of Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0-12</td>
<td>214</td>
<td>32</td>
<td>198</td>
<td>32</td>
<td>0.84</td>
<td>0.84</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>CD4&lt;200</td>
<td>12-24</td>
<td>150</td>
<td>23</td>
<td>138.5</td>
<td>25</td>
<td>0.82</td>
<td>0.69</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td>24-36</td>
<td>102</td>
<td>10</td>
<td>97</td>
<td>10</td>
<td>0.9</td>
<td>0.62</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>36-48</td>
<td>82</td>
<td>11</td>
<td>76.5</td>
<td>4</td>
<td>0.95</td>
<td>0.58</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>48-60</td>
<td>67</td>
<td>3</td>
<td>65.5</td>
<td>3</td>
<td>0.95</td>
<td>0.56</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0-12</td>
<td>120</td>
<td>22</td>
<td>109</td>
<td>5</td>
<td>0.95</td>
<td>0.95</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>CD4≥200</td>
<td>12-24</td>
<td>93</td>
<td>11</td>
<td>87.5</td>
<td>3</td>
<td>0.97</td>
<td>0.92</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td>24-36</td>
<td>79</td>
<td>6</td>
<td>76</td>
<td>5</td>
<td>0.93</td>
<td>0.86</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>36-48</td>
<td>68</td>
<td>8</td>
<td>64</td>
<td>2</td>
<td>0.97</td>
<td>0.83</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>48-60</td>
<td>58</td>
<td>5</td>
<td>55.5</td>
<td>2</td>
<td>0.96</td>
<td>0.80</td>
<td>0.04</td>
<td></td>
</tr>
</tbody>
</table>

The Table-2.2 shows that the survival probabilities of the years of 1,2,3,4 and 5 of the patients according to CD4 less than 200 were 0.84, 0.69, 0.62, 0.58, 0.56 whereas CD4 greater than 200 for the first year were 0.95, 0.92, 0.86, 0.83, 0.80 respectively. It shows that the survival probability for the CD4 greater than 200 is higher than the survival probabilities of less than 200 counts.
Table-2.3 Mean survival times for Gender

<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>Female</td>
<td>43.369</td>
<td>2.837</td>
<td>37.809</td>
</tr>
<tr>
<td>Male</td>
<td>46.519</td>
<td>1.405</td>
<td>43.766</td>
</tr>
</tbody>
</table>

Survival Functions

Figure-2.1 Survival curves for HIV-TB patients according to Gender

Table-2.4 Nonparametric test for gender

<table>
<thead>
<tr>
<th>Log Rank (Mantel-Cox)</th>
<th>Chi-square</th>
<th>Df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breslow (Generalized Wilcoxon)</td>
<td>1.080</td>
<td>1</td>
<td>0.299</td>
</tr>
<tr>
<td>Tarone-Ware</td>
<td>0.850</td>
<td>1</td>
<td>0.356</td>
</tr>
</tbody>
</table>

Fig-2.1 shows the survival curves for comparing of male and female alone, and the table-2.3 shows the mean survival time for male is 46.519 (s.e. 1.045), and female 43.369 (s.e. 2.837). It indicates that the survival time for the event for male is higher than the female. However there is no significance difference between these two groups in non-parametric tests.
Table-2.5  Mean survival times for Treatment

<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>Reg 6M</td>
<td>46.600</td>
<td>1.744</td>
<td>43.183</td>
</tr>
<tr>
<td>Reg 9M</td>
<td>44.970</td>
<td>1.834</td>
<td>41.376</td>
</tr>
</tbody>
</table>

Survival Functions

Figure-2.2 Survival curves for Treatments HIV-TB according to Treatments

Table-2.6 Nonparametric Test for Treatments

<table>
<thead>
<tr>
<th>Test</th>
<th>Chi-square</th>
<th>Df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Rank (Mantel-Cox)</td>
<td>0.359</td>
<td>1</td>
<td>0.549</td>
</tr>
<tr>
<td>Breslow (Generalized Wilcoxon)</td>
<td>0.589</td>
<td>1</td>
<td>0.443</td>
</tr>
<tr>
<td>Tarone-Ware</td>
<td>0.475</td>
<td>1</td>
<td>0.491</td>
</tr>
</tbody>
</table>

Fig.2.2 shows the survival curves for comparing of the 6M & 9M treatments alone, the table2.5 shows the 6 month mean survival time is 46.600 (s.e. 1.744), for 9 month is 44.970 (s.e. 1.834). It shows that the survival time for 6 month is little higher than 9 month treatment and it shows that there is no significance between the treatments.
Table-2.7 Mean Survival times for CD4 counts

<table>
<thead>
<tr>
<th>Covariate</th>
<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></td>
<td>Lower Bound</td>
</tr>
<tr>
<td>CD4&lt;200</td>
<td>Female</td>
<td>37.741</td>
<td>3.095</td>
<td>30.087</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>42.767</td>
<td>1.876</td>
<td>39.090</td>
</tr>
<tr>
<td>CD4≥200</td>
<td>Female</td>
<td>51.668</td>
<td>3.440</td>
<td>44.925</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>54.059</td>
<td>1.659</td>
<td>50.807</td>
</tr>
</tbody>
</table>

Figure-2.3 Survival curves for HIV-TB female

Figure-2.4 Survival curves for HIV-TB male
Table 2.8 Nonparametric Test for CD4

<table>
<thead>
<tr>
<th>Sex_Code</th>
<th>Chi-square</th>
<th>Df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log Rank (Mantel-Cox)</td>
<td>4.868</td>
<td>1</td>
<td>0.027</td>
</tr>
<tr>
<td>Breslow (Generalized Wilcoxon)</td>
<td>4.123</td>
<td>1</td>
<td>0.042</td>
</tr>
<tr>
<td>Tarone-Ware</td>
<td>4.616</td>
<td>1</td>
<td>0.032</td>
</tr>
<tr>
<td>Male</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log Rank (Mantel-Cox)</td>
<td>13.265</td>
<td>1</td>
<td>0.000</td>
</tr>
<tr>
<td>Breslow (Generalized Wilcoxon)</td>
<td>13.835</td>
<td>1</td>
<td>0.000</td>
</tr>
<tr>
<td>Tarone-Ware</td>
<td>13.877</td>
<td>1</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Fig. 2.3 and 2.4 shows the survival curves for the comparisons of male and female survival curves for CD4 counts and Table 2.7 shows the mean survival time for the event of interest of CD4 < 200 for male is 42.767 (s.e. 1.876), female is 37.741 (s.e. 3.905) and for CD4 ≥ 200 count for male is 54.059 (s.e. 1.659), female is 51.668 (s.e. 5.440). The mean survival time for CD4 counts ≥ 200 is higher than CD4 < 200 counts in both sex. All the three non-parametric tests shows significance difference between CD4 < 200 & CD4 ≥ 200 and among the male and female patients.
Table 2.9: Mean survival times for Treatment

<table>
<thead>
<tr>
<th>Covariate</th>
<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></td>
<td>Upper Bound</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower Bound</td>
</tr>
<tr>
<td>Reg_6M</td>
<td>Female</td>
<td>46.169</td>
<td>3.943</td>
<td>53.897</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>46.674</td>
<td>1.944</td>
<td>50.485</td>
</tr>
<tr>
<td>Reg_9M</td>
<td>Female</td>
<td>40.780</td>
<td>4.018</td>
<td>48.654</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>46.285</td>
<td>2.038</td>
<td>50.280</td>
</tr>
</tbody>
</table>

Figure 2.5: Survival curves for HIV-TB female (Regimen)
Fig. 2.5 and 2.6 shows the survival curves for Comparisons of Male and Female Survival Curves for treatments and table 2. 9 shows that the mean survival time for the event of interest for 6 months treatment for male 46.674 (s.e 1.944) and for female 46.169 (s.e. 3.943). For 9 months treatment with respect to male is 46.285 (s.e. 2.038) and for female 40.78 (s.e. 4.018). The mean survival time for 9 months and 6 months treatments is same in male whereas in female 6 months treatment is higher than the 9 month treatment. However the Log-Rank (Mantel-Cox), Breslow (Generalized Wilcoxon) and Tarone-Ware tests shows non significance results in both sex.
| Covariates | Hazard Ratio | Std. Err. | z    | P>|z| | 95% Conf. Interval |
|------------|-------------|-----------|------|-----|---------------------|
| Age        | 1.010677    | 0.014469  | 0.74 | 0.458 | 0.982711 - 1.039438 |
| Weight     | 0.965016*   | 0.013629  | -2.52| 0.012 | 0.93867 - 0.992101  |
| Sex        | 0.946858    | 0.250485  | -0.21| 0.836 | 0.563773 - 1.590248 |
| Reg.       | 1.129673    | 0.238033  | 0.58 | 0.563 | 0.747475 - 1.707296 |
| CD4        | 0.348185*   | 0.094013  | -3.91| 0.000 | 0.205106 - 0.591074 |

-2LL = 956.596

*P<0.05

Table-2.11 shows the Cox model containing five covariates regimen, sex, weight, CD4 counts and age. It gives the regression estimates using the proportional hazards model for the event of death. The hazard ratios show almost similar values for all the covariates except the CD4 counts. The hazard ratio for CD4 is 0.35. The covariates weight and CD4 shows significant values and the deviance value is 956.596.
Table-2.12 Extended Cox Model Adjusting age

| Covariates | Haz. Ratio | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|------------|------------|-----------|-------|-----|-------------------|
| T_Age      | 1.000107   | 0.000667  | 0.16  | 0.873| 0.998801 - 1.001414 |
| Weight     | 0.96608*   | 0.013624  | -2.45 | 0.014| 0.939744 - 0.993154 |
| Sex        | 0.976831   | 0.25522   | -0.09 | 0.929| 0.58536 - 1.630105 |
| Reg        | 1.12522    | 0.237297  | 0.56  | 0.576| 0.744266 - 1.701166 |
| CD4        | 0.346324*  | 0.09351   | -3.93 | 0.000| 0.20401 - 0.587913 |

-2LL = 957.1089

*P<0.05

Table-2.12 shows for extended Cox PH model for the time dependent covariate age adjusted with other covariates regimen, sex, weight, and CD4 counts shows almost the same results as Cox PH model. The difference is this model shows higher deviance value 957.1089 compared to the previous model.

Table-2.13 Extended Cox model adjusting weight

| Covariates | Haz. Ratio | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|------------|------------|-----------|-------|-----|-------------------|
| T_Wt       | 0.999249   | 0.000587  | -1.28 | 0.201| 0.998099 - 1.0004 |
| Age        | 1.008036   | 0.014558  | 0.55  | 0.579| 0.979903 - 1.036976 |
| Sex_Code   | 0.824302   | 0.212534  | -0.75 | 0.454| 0.497298 - 1.366332 |
| Reg_Code   | 1.110261   | 0.233934  | 0.50  | 0.620| 0.734641 - 1.677933 |
| CD4_0m     | 0.340716*  | 0.091911  | -3.99 | 0.000| 0.200805 - 0.578111 |

-2LL = 961.6547

*P<0.05

Table-2.13 shows for extended Cox PH model for the time dependent covariate weight adjusted with other covariates regimen, sex, age, and CD4 counts also shows almost the same results as Cox PH model and extend PH model for age. The deviance value for this model is 961.6547. The deviance of Cox PH model 956.596 is less compared with the two extended models for the time dependent covariates weight and age. It shows that Cox PH model fits better than the extended Cox PH models. The covariates weight and CD4 shows the significance values.
### Table-2.14 Exponential Model

<table>
<thead>
<tr>
<th>Covariates</th>
<th>Hazard Ratio</th>
<th>Std. Err.</th>
<th>Wald</th>
<th>P-value</th>
<th>95% Conf. Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>1.011488</td>
<td>0.014825</td>
<td>0.78</td>
<td>0.436</td>
<td>0.982846 1.040965</td>
</tr>
<tr>
<td>Weight</td>
<td>0.963840</td>
<td>0.012707</td>
<td>-2.79</td>
<td>0.005</td>
<td>0.939253 0.989070</td>
</tr>
<tr>
<td>cd4</td>
<td>0.995788</td>
<td>0.000934</td>
<td>-4.50</td>
<td>0.000</td>
<td>0.993959 0.997621</td>
</tr>
</tbody>
</table>

-2LL = 557.2415  
LR = 40.00

### Table-2.15 Weibull Model

<table>
<thead>
<tr>
<th>Covariate</th>
<th>β value</th>
<th>Std. Err.</th>
<th>Wald</th>
<th>P-value</th>
<th>95% Conf. Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.01058</td>
<td>0.014619</td>
<td>0.72</td>
<td>0.469</td>
<td>-0.01807 0.039233</td>
</tr>
<tr>
<td>Weight</td>
<td>-0.03548</td>
<td>0.013203</td>
<td>-2.69</td>
<td>0.007</td>
<td>-0.06136 -0.00961</td>
</tr>
<tr>
<td>cd4</td>
<td>-0.00413</td>
<td>0.000937</td>
<td>-4.40</td>
<td>0.000</td>
<td>-0.00596 -0.00229</td>
</tr>
<tr>
<td>_cons</td>
<td>-2.43823</td>
<td>0.712865</td>
<td>-3.42</td>
<td>0.001</td>
<td>-3.83541 -1.04104</td>
</tr>
<tr>
<td>/ln_p</td>
<td>-0.11313</td>
<td>0.089354</td>
<td>-1.27</td>
<td>0.205</td>
<td>-0.28826 0.062</td>
</tr>
<tr>
<td>p</td>
<td>0.893035</td>
<td>0.079796</td>
<td></td>
<td></td>
<td>0.749567 1.063962</td>
</tr>
<tr>
<td>1/p</td>
<td>1.119777</td>
<td>0.100056</td>
<td></td>
<td></td>
<td>0.939883 1.334104</td>
</tr>
</tbody>
</table>

-2LL = 555.558  
LR = 37.4

### Table-2.16 Log-Normal Model

<table>
<thead>
<tr>
<th>Covariate</th>
<th>β value</th>
<th>Std. Err.</th>
<th>Wald</th>
<th>P-value</th>
<th>95% Conf. Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>-0.01842</td>
<td>0.016254</td>
<td>-1.13</td>
<td>0.257</td>
<td>-0.05028 0.013435</td>
</tr>
<tr>
<td>Weight</td>
<td>0.046109</td>
<td>0.015587</td>
<td>2.96</td>
<td>0.003</td>
<td>0.015558 0.07666</td>
</tr>
<tr>
<td>cd4</td>
<td>0.003983</td>
<td>0.000919</td>
<td>4.34</td>
<td>0.000</td>
<td>0.002183 0.005784</td>
</tr>
<tr>
<td>_cons</td>
<td>2.427595</td>
<td>0.797111</td>
<td>3.05</td>
<td>0.002</td>
<td>0.865287 3.989903</td>
</tr>
<tr>
<td>/ln_sig</td>
<td>0.532257</td>
<td>0.081731</td>
<td>6.51</td>
<td>0.000</td>
<td>0.372067 0.692447</td>
</tr>
<tr>
<td>sigma</td>
<td>1.702772</td>
<td>0.139169</td>
<td></td>
<td></td>
<td>1.450731 1.9986</td>
</tr>
</tbody>
</table>

-2LL = 545.125
The above AFT model tables show the analysis of different parametric distributions Exponential, Weibull, Log normal and Log logistic. For these models, the continuous covariates age weight and CD4 were considered. In all the four models weight and CD4 shows significant values. The deviance for lognormal model is 545.125 which is less compared to the all other three models which implies that the data fits better in the lognormal model.

### 2.9 Summary

The survival function can be estimated by methods like actuarial table (Life), Kaplan-Meier and compared by Logrank, Tarone-Ware and Breslow test. The actuarial method can describe the survival times of homogeneous subjects when the survival times subject to censoring the assumptions of independence of risk of events and withdrawal must hold. The actuarial estimates are obtained by divide in the time period into a series of time intervals and the numbers of individuals events are censored in each interval are recorded. It is appropriate only when the sample size is fairly large.
Kaplan and Meier’s (1958) product limit method is the most commonly used technique for estimating the survivorship function for samples of small, moderate and even for large sizes. It is also known as product limit method based on conditional probabilities. Researchers may wish to compare the survival curves of two different groups. The intuition of Kaplan-Meier estimation is the idea of a chronological accumulation of probabilities. The log rank test is a large sample chi-square test that uses as its test criterion or stochastic that provides an overall comparison of the Kaplan and Meier being compared. The Log-rank test assigns equal weight to each observation. The log rank statistic makes use of observed versus expected cell counts over categories of outcomes. It is to test equality of survival functions by weighting all time points the same. The categories of log rank tests are defined by each of the ordered failure times for the entire set of data being analyzed. The statistics summarizes the extent to which the observed survival times in the two groups of data from those expected under the null hypothesis of no differences between groups. Gehan test assigns greater weight to the earlier observations. Its formulation derives from a procedure known as the Wilcoxon rank sum test for the nonparametric comparison of two groups. Tarone-Ware test, in its weighing scheme, is intermediate between the Log-rank test and Gehan test. Breslow test is the Gehan test. It is appropriate when the number of groups being compared is greater than 2. The outcome of nonparametric tests and Kaplan-Meier curve of this data shows that there was no significance difference in the treatment groups. Though the nonparametric tests were shown non significant for the treatment groups, the survival curves for the treatments, there is little better survival times for the patients who had 6 months treatment. The mean survival estimate for 9 months treatment is 44.970 and the 6 months treatment survival estimate of 46.600. The mean survival estimate for female is 43.369 which is
lesser than the male survival estimate of 46.519, but it shows no significance
difference between these two groups. The survival curves for the treatments are also
shown some better survival times for the patients who had 6 months treatment in both
sex groups. For the covariate CD4 counts, all the non parametric tests Log rank,
Breslow and Tarone-Ware shows significance results. The survival time for male who
had CD4 counts <200 is 37.741 which is less compared to the survival time for
patients who had Cd4 ≥ 200 counts of 51.668 and for female the survival time for male
who had CD4 counts <200 is 42.767 which is less compared to the survival time for
patients who had Cd4 ≥ 200 counts of 54.059. The survival curves for the treatments
are also shown some better survival times for the patients who had ≥ 200 counts.

The Kaplan Meier curves and its estimate of times showed the survival
experience for the events occurring for two treatment groups. All the three non
parametric tests show that non significance results for the covariates gender,
treatments. The covariate CD4 counts with respect to male and female shows
significant results. The 6months treatment shows better survival time compared to
9months treatment and the mean survival estimates of male had higher survival times
compared to female patients. The same results reflect in life table and its survival
curves.

The techniques of Cox proportional hazard model development is about
modeling time to event and its relationship to a set of the one or more explanatory
variables in the presence of censoring. The primary reason is that the Cox PH model is
robust for many survival analysis situations. The ML estimators of the parameters in
the Cox model is presented and pointed out that the ML procedure maximizes a partial
likelihood that focuses on probabilities at failure (event) times only. We have also
presented a general formula for estimating a hazard ratio that compared two
specifications of the X’s and applied for comparing two exposure groups adjusted for other variables of HIV-infected tuberculosis trial. The adjusted survival curves for comparing different groups are discussed. The extended Cox model that can be used to evaluate PH assumption for a given predictor adjusted predictors already satisfying the PH assumption is presented. The specific form of extended model will approximate the analysis for given a survival data involving one or more time dependent variables and heavy side function for checking PH assumptions. Anderson et al., (1993) noted that the underlying hazard of the fitted Cox regression models varied so regularly that a hypothesis of Weibull underling intensity should be acceptable.

The deviance of Cox PH model 956.596 is less compared with the two extended models for the time dependent covariates weight and age. It shows that Cox PH model fits better than the extended Cox PH models.