3.0 INTRODUCTION

The use of parametric models is deeply entrenched in statistical practice. The reason for this is that they provide an approximate description of a complete set of data by means of a qualitative information and very few simple quantities (the approximations or estimates of the parameters of the model) and that they together with the parameter values, provide a complete and easily described stochastic model for the generation of the observed data and other future or future observations. They thus fulfill one of the main aims of statistics namely data reduction or data condensation and they also allow the application of the methods of probability theory to the complete description of the whole data set. In other words, parametric models allow the separation of the full information of a data set into pure structure and pure random variability.

The development of estimation and hypothesis testing procedures for survival models was one of the major areas of activity in 1960's. Books by many authors (e.g., Bain, 1978; Gross and Clark, 1975; Lawless, 1982; Mann et al., 1974 and Nelson, 1982) review much of the work in this area. Before 1959, a considerable amount of work has been done on inference procedures for the exponential distribution with uncensored and censored data (Epstein and Sobel, 1953, 1954; Bartholomew, 1957) and for the
normal and lognormal distributions. However relatively little had been done with the methodology for censored data in general and for distribution such as the Weibull, gamma and lognormal with censored data. Inference procedures for univariate life distributions such as the gamma (e.g., Engelhart and Bain, 1978), the generalised gamma (e.g., Farewell and Prentice, 1977; Lawless, 1982), the lognormal (e.g., Nelson and Schemer, 1979) and the inverse Gaussian (e.g., Chikkara and Folks, 1977) have been done in recent years. In addition, our understanding of general large sample maximum likelihood methods for the use with censored data has improved considerably. An early paper by Halperin (1952) addressed large-sample properties of maximum likelihood for the case in which the data is Type II censored. Later works by Blight (1970), Cox (1975), Aalen (1978) and Kalbfleisch and Prentice (1980) motivated works in other types of censoring.

When fitting parametric survival functions, the choice of the form of the survival function is supposed to have already been decided. Our interest is then in estimating values of the parameters appearing in the mathematical form for survival function of the family of distributions considered.

We distinguish two broad groups of methods for fitting distributions: graphical and analytical. Most graphical methods rely on plotting some functions of the hazard rate or cumulative hazard function against some function of t. The functions are so chosen that if the parametric form of the survival function is reasonably appropriate, an approximately straight line plot will
The constant hazard, \( h(t) = \lambda \), characterises the exponential family, with density \( f(t, \lambda) = \lambda \exp(-\lambda t) \) and the survival function \( S(t, \lambda) = \exp(-\lambda t) \). Radioactive decay times are a classic example of data that follow an exponential distribution. It would be reasonable to assume an exponential death density if the cause of death occurs according to a Poisson process with a constant rate. If for example, an individual is subjected to random events such as a blood clot or thrombosis that causes the body to fail if and only if that event occurs, we would expect the exponential death density to govern the length of life of the individual. Although the assumption of a constant hazard is too restrictive for most of the biomedical applications, the exponential provides a convenient starting point for analysis and a useful standard of comparison for assessing the behaviour of other distributions. The exponential distribution was the first survival model to become widely used, partly due to its attractive computational features. However, the theoretical investigations revealed by Zelen and Dannemiller (1961), that these methods are very sensitive to departure from the exponential model and consequently should be applied with caution. The appropriateness of the exponential model for a given set of survival data may be checked by plotting \( \hat{A}(t) \) or equivalently \( -\log(\hat{S}(t)) \) versus \( t \); such a plot should approximate a straightline through the origin.

The behaviour of the hazard function is related to the relative dispersion of the distribution in the sense that distributions with hazard functions that are strictly increasing,
be obtained. The plotting can be facilitated by the use of special graph paper—probability paper or hazard paper. Especially useful in survival data analysis are cumulative hazard papers on which \( t, h(t) \) can be entered directly to give a straight line plot (Nelson, 1972). Although graphical methods give reasonably useful results in many cases, they are often subjective in nature making one to prefer an analytical approach.

Among analytical methods, we distinguish ad hoc methods and standard methods. The method of percentiles, based on making the fitted survival function equal to the observed survival function, at a few selected points, is an ad hoc method. On the other hand, methods such as maximum likelihood, minimum chi-square or least squares are regarded as mathematically and statistically sound, although the soundness depends on the accuracy of the assumption on which they are based.

3.1 EXPLORING SPECIAL PARAMETRIC MODELS

It will be occasionally possible to specify the reference distribution exactly. But more often the form of the reference distribution can be given only in term of a given number of unknown parameters, which must thus and be estimated from the data. The choice of a family in any specific application may be governed by a biological theory, for example, multistage models for carcinogenesis lead naturally to Weibull distribution for the time to tumour induction, and the two hit model in bio-assay yields a particular gamma distribution.
constant or strictly decreasing, have co-efficients of variation that are respectively, less than, equal to and greater than unity. It is helpful if the density, survival and hazard functions can all be expressed in closed forms. Otherwise, likelihood functions from censored data cannot be obtained explicitly.

The Weibull distribution is probably the most widely used parametric survival model in biomedical as well as technical applications. The Weibull model provides a fairly flexible class of distributions and includes the exponential distribution as a special case. The hazard and survival functions have the form

\[ h(t) = \lambda p (\lambda t)^{p-1} \]  \hspace{1cm} [3.1.1]

\[ S(t) = \exp \left( \left( -\lambda t \right)^p \right) \hspace{1cm} t \geq 0 \]  \hspace{1cm} [3.1.2]

where \( \lambda > 0 \) is the inverse scale parameter and \( p > 0 \) is a shape parameter. The hazard function is monotonically increasing if \( p > 1 \) and monotonically decreasing if \( 0 < p < 1 \) and constant if \( p = 1 \). The Weibull distribution appears as one of the asymptotic distributions of the smallest extreme value distribution and this fact motivates its use in certain applications. From the above it follows that

\[ \log \Lambda(t) = p \log \lambda + p \log t \]  \hspace{1cm} [3.1.3]

and the appropriateness of the Weibull model can then be checked by plotting \( \log \hat{\Lambda}(t) \) verses \( \log t \).

The gamma model provides an alternative two parameter family of distributions that includes' exponential as a special case. The gamma distribution has density function of the form

\[ f(t) = \lambda^\alpha t^{\alpha-1} \exp(-\lambda t) / \Gamma(\alpha) \hspace{1cm} t \geq 0 \]  \hspace{1cm} [3.1.4]

where \( \lambda > 0 \) is an inverse scale parameter and \( \alpha > 0 \) is a shape
parameter. The survival function can be expressed as

$$S(t) = 1 - I(\alpha, \lambda t)$$  \hspace{1cm} (3.1.5)

where \( I(\alpha, x) \) is the incomplete gamma function

$$I(\alpha, x) = \int_{0}^{x} u^{\alpha-1} e^{-u} du$$  \hspace{1cm} (3.1.6)

Closed forms expressions are only available for integer values of \( \alpha \) and the gamma model is therefore less attractive than Weibull model. The survival function is monotonically increasing if \( \alpha > 0 \) and monotonically decreasing if \( 0 < \alpha < 1 \). For \( \alpha = 1 \), the hazard is constant. For gamma model plots of quantiles based on parametric assumption against quantiles based on product limit estimates give a straight line if the models are appropriate.

An application of the gamma density to biomedical situation can be seen from an example. Suppose we are studying a group of patients suffering from a kidney disease in which the failure rate of each kidney is constant and equal to \( \lambda \). For a patient to die from the disease, both kidneys must fail. Hence the hazard rate of these patients is \( \lambda \left[ \lambda t/(1+\lambda t) \right] \) which is the hazard rate of a gamma density. The hazard rate is an increasing function of time for \( \lambda > 0 \).

The lognormal distribution has also been widely used in survival analysis. Like gamma model, no closed form expressions are available for \( S(t) \) and \( h(t) \). The distribution is most easily specified through \( \log T \) having a normal distribution with mean \( \mu \) and variance \( \sigma^2 \). With this parameterization, the lognormal survival function is
\[ S(t) = 1 - \frac{\phi((\log(t) - \mu) / \sigma)}{} \]  

where \( \phi \) is the distribution function of standard normal distribution. The plot of \( S(t) \) against log \( t \) gives a linear plot if the lognormal model is appropriate. The lognormal hazard function has the value 0 at \( t=0 \), increases to a maximum and then decreases with a limiting value of 0 as \( t \) tends to infinity. The behaviour is unattractive in some of the medical applications.

The generalised gamma model discussed in Chapter II provides a flexible three parameter family of distributions that includes the preceding four models as special cases. Other useful parametric models for which \( \lambda(t) \) or \( \log \lambda(t) \) is a polynomial of low order are, the log-logistic model (Kalbfleisch and Prentice, 1980) the inverse Gaussian model (Chhikara and Folks, 1977), Gompertz and the piecewise exponential models. If we plot \( \log[-\log S(t)] \) against \( t \) we obtain an approximately straight line graph when Gompertz model is appropriate. Similarly for the Mekam-Gompertz \( \log[h(t) - A] \) against \( t \) gives a straight line for properly chosen \( A \).

Another useful family of families is obtained by allowing the parameter \( \lambda \) in the exponential distribution itself to be a random variable. Thus, for example, if \( \lambda \) has a gamma distribution and the conditional distribution of \( T \) given \( \lambda \) is exponential, then unconditionally \( T \) will have a Pareto distribution. All such compound exponential distributions must have decreasing hazards.
All parametric models can be modified to allow for an initial event-free period by introduction of a threshold parameter \( \delta \) by replacing \( T \) by \( T' = T - \delta \).

Plotted data sometimes indicate that it is not possible to fit a single distribution function over the whole range of data. In studies of tuberculosis trials, data often indicate rather slow response rate in the first few weeks after the start of treatment, then very high response rate during some period of time, but after certain critical period, response rate is very low. In general, there may be more than two such periods. The points of such periods are often suggested by graphical displays; even if no additional information on the disease is available. If we denote the survival functions in the successive periods by \( S_0(t), S_1(t), \ldots; S_i(t) \) and \( S_j(t) \) \((i \neq j)\) can even belong to different families; in practice, they are usually different members of the same family.

Maximum likelihood estimation and large-sample likelihood methods are the inference procedures generally used, as the presence of censoring makes exact distributional results extremely complicated in most situations. In the case of no censoring or Type II censoring alternative procedures are available for the exponential and the Weibull models. A review of these methods is given by Lawless (1982). An unified approach to the asymptotic theory of maximum likelihood estimation for the parametric survival models with censoring has been given by Borgan (1984).
3.2 ANALYSIS OF SURVIVAL DATA: PARAMETRIC MODELS WITH COVARIATES.

In studies of patients with tuberculosis and other diseases, it has long been recognised that certain characteristics of the patient may markedly influence survival, but only in the past two decades, powerful and flexible statistical methods have been available for incorporating these co-variates, often referred to as prognostic factors in the analysis of survival data. Most earlier works evaluated prognostic factors by constructing separate survival curves for different categories of patients. Methods were not available for dealing with more than one variable at a time except by further sub-categorization and construction of separate curves, although some work described earlier the use of parametric survival models based on exponential distribution (e.g., Feigl and Zelen, 1965). The most important development was the proportional hazard semi-parametric model by Cox (1972). This is further considered in Chapter V. In this section, we shall not attempt a general review of the literature on survival models incorporating covariate information, which by now has become rather extensive, but will instead concentrate on use of the Weibull and other models in actual analysis of tuberculosis data.

Before discussing these models it is appropriate to review our reasons for studying prognostic factors. One important reason is to understand how a disease behaves. Is the prognosis similar in men and in women? Is age an important prognostic factor? Does the extent of disease and its bacteriological grade
influence the outcome and are the values of certain laboratory tests and the results of physical examination significantly correlated with length of survival? If several of these variables are important, how do they act together? When these questions have been answered, we can use the results of our statistical analysis to predict survival for groups of patients, a second important goal. A third goal is to enable us to perform adjusted survival analysis when comparing group of patients for which imbalances exist in the distribution of important prognostic factors. This need arises especially in the analysis of retrospectively collected data, but such analysis can also be important in interpreting data from randomized clinical trials studying subsets of patients. A fourth goal of analyzing prognostic factors is to aid us in the design of new studies. Our analysis may tell us how long to expect the patients to live and how to stratify the randomization in such a way as to assure balance on the most important prognostic factors.

Although it is possible to miss important variables by studying them one at a time (because their effects may be confounded by other variables), it is usually impractical to try to fit a multivariate model using all candidate variables without some preliminary screening unless the number of variables under study is small, say 15 or fewer. We therefore identify a group of potentially prognostic variables by using prior information, constructing survival curves for various categories of patients, or employing the methods described in the previous sections. The
next step is to choose a multivariate survival model in order to evaluate the effects on these variables when they are studied simultaneously. Applications to tuberculosis survival data are considered in section 3.5.

3.3 ESTIMATION OF EXPONENTIAL SURVIVAL PROBABILITIES WITH CONCOMITANT VARIABLES

In simple linear or multiple linear regression model, estimation of the regression parameters is accomplished by ordinary least squares procedures. In analysing patients survival data in the presence of concomitant information, some modifications are necessary. In this section, we consider the statistical model put forth by Feigl and Zelen (1965) for one concomitant variate and discussed by Gross and Clark (1975).

Suppose $n$ patients are on study and the survival density function for the $i$-th patient is

$$f_i(t_i) = \lambda_i \exp(-\lambda_i t_i); \lambda_i > 0, t_i \geq 0, i=1,..,n$$  \[3.3.1\]

Initially we assume that all patients are followed until the event of interest (death or response or toxicity). Further more, let $Z_i$, be the observed values of a concomitant variable or covariate, such that

$$E(t_i) = \frac{1}{\lambda_i} = a + bZ_i \quad i = 1, 2, \ldots n.$$  \[3.3.2\]

Thus the mean survival time of patients is assumed to be linearly related to the concomitant variable. For example, if $Z_i$ represents mantoux levels of patients with pulmonary disease, we might expect small values of the concomitant variate to correspond to relatively small values of the mean survival time of the disease.
Conversely, we should expect large values of the concomitant variates to correspond to large values of the mean survival time of the disease. \( Z_1 \) could also refer to a transformed value. To estimate the parameters \( a \) and \( b \) we use the maximum likelihood. Let \( L(a,b) \) be the likelihood function for \( a \) and \( b \), given the observations \((Z_1, t_1) \ldots (Z_n, t_n)\). Then

\[
L(a,b) = \prod (a+bZ_i)^{-1} \exp \left(-\frac{(a+bZ_i)^{-1} t_i}{a+bZ_i} \right)
\]

Then, differentiating \( \log L \) with reference to \( a \) and \( b \), respectively, the maximum likelihood estimators of \( a \) and \( b \) are obtained by usual methods such as Newton-Raphson iterative technique. The initial estimates \( a_0 \) and \( b_0 \) can be obtained by plotting \( t_i \) against \( Z_i \) and estimating from it approximately or computing with the regression equation

\[
E(t_i) = a_0 + b_0 Z_i \quad i = 1, \ldots, n.
\]

when the least square estimates \( \hat{a}_0 \) and \( \hat{b}_0 \) obtained from the above equations become the initial estimates in the iterative procedure.

If \( \hat{a}_k \) and \( \hat{b}_k \) denote the estimates at the \( k \)-th iteration, the values for the \((k+1)\)th iteration are obtained by solving the two simultaneous equations (3.3.5) and (3.3.6) in the two unknowns \( \delta a_k \) and \( \delta b_k \) namely

\[
A_k \delta \hat{a}_k + B_k \delta \hat{b}_k = D_k
\]

and

\[
B_k \delta \hat{a}_k + C_k \delta \hat{b}_k = E_k
\]

where

\[
A_k = \frac{\delta^2}{\delta a_k^2} (\log L) \\
B_k = \frac{\delta^2}{\delta a_k \delta b_k} (\log L) \\
C_k = \frac{\delta^2}{\delta b_k^2} (\log L) \\
D_k = \frac{\partial}{\partial a_k} (\log L) \\
E_k = \frac{\partial}{\partial b_k} (\log L)
\]

and

\[
\delta a_k = a_{k+1} - a_k \\
\delta b_k = b_{k+1} - b_k
\]

[3.3.9]
The asymptotic variance-covariance matrix for \((a, b)\) is obtained from

\[
\begin{bmatrix}
\text{Var}(\hat{a}) & \text{Cov}(\hat{a}, \hat{b}) \\
\text{Cov}(\hat{a}, \hat{b}) & \text{Var}(\hat{b})
\end{bmatrix} = - \frac{\delta^2 \log L}{\delta a \delta b}
\begin{bmatrix}
\mathbf{E}(\frac{\delta^2 \log L}{\delta a^2}) & \mathbf{E}(\frac{\delta^2 \log L}{\delta a \delta b}) \\
\mathbf{E}(\frac{\delta^2 \log L}{\delta a \delta b}) & \mathbf{E}(\frac{\delta^2 \log L}{\delta b^2})
\end{bmatrix}
\]

where

\[
\text{Var}(\hat{a}) = \Delta^{-1} \sum_{1}^{n} z_1^2 r_1
\]

\[
\text{Var}(\hat{b}) = \Delta^{-1} \sum_{1}^{n} r_1
\]

\[
\text{Cov}(\hat{a}, \hat{b}) = \Delta^{-1} \sum_{1}^{n} z_1 (r_1)
\]

\[
\Delta = \left( \sum_{1}^{n} z_1^2 r_1 \right) \left[ \sum_{1}^{n} r_1 \right] - \left( \sum_{1}^{n} z_1 \right) (r_1)^2
\]

\[
R_1 = (a + b z_1)^{-2}
\]

Now we have

\[
\text{Var}(\hat{a} + \hat{b} z_1) = \text{Var}(\hat{a}) + z_1^2 \text{Var}(\hat{b}) + 2 z_1 \text{Cov}(\hat{a}, \hat{b})
\]

using (3.3.12) a 100 \((1-\alpha)\) percent confidence interval for \((a + b z_1)\), \(i = 1, 2, \ldots, n\) is approximately

\[
(a + \hat{b} z_1) - Z(1-\alpha/2) \hat{\sigma} < a + b z_1 < (a + \hat{b} z_1) + Z(1-\alpha/2) \hat{\sigma}
\]

where

\[
\hat{\sigma} = \sqrt{\text{Var}(\hat{a} + \hat{b} z_1)}
\]

Thus a 100 \((1-\alpha)\) percent simultaneous confidence interval for \(S_i(t) = \exp(-t(a + b z_i))\), \(i = 1, 2, \ldots, n\), the probabilities that each of the \(n\) patients survive to at least time \(t\), is

\[
\exp(-t c_{11}^{-1}) \leq S_i(t) \leq \exp(-t c_{21}^{-1})
\]

where

\[
c_{11} = a + \hat{b} z_1 - Z(1-\alpha/2) \hat{\sigma}
\]

\[
c_{21} = a + \hat{b} z_1 + Z(1-\alpha/2) \hat{\sigma}
\]

To test the goodness of the fitted model, we compare the expected
and observed numbers of patients falling (dying) in certain predetermined time intervals, using a chi-square goodness of fit test. If the time axis is divided into \( k \) sub intervals, these determine for the \( i \)-th patient the \( k \)-th quantile intervals.

\[
0 < t_1 < t_1(1/k), \quad t_1(1/k) \leq t_1 < t_1(2/k), \ldots, t_1((k-1)/k) \leq t_1 < \infty
\]

The probability is \( 1/k \) the \( i \)-th patient dies in any one interval, \( i=1,2,\ldots,k \). If the model is a good fit, we would expect the survival times to be equally distributed in the \( k \) intervals.

Zippin and Armitage (1966) modify the Feigl-Zelen model when the survival data are censored. Specifically, suppose \( n \) patients are on study in a clinical trial setting with \((3.3.1)\) as the density function for the \( i \)-th patient. However, the \( i \)-th patient survival time \( t_i \) is known only if \( t_i \leq T_1 \), where \( T_1 \) is the maximum observation time of the \( i \)-th patient, \( i=1,2,\ldots,n \).

It follows that

\[
L(a,b) = \prod_{i=1}^{n} \left[ (a + bZ_i)^{-1} \exp(- (a + bZ_i)^{-1} t_i \delta_i) \right] \left[ \exp \left( - (a + bZ_i)^{-1} t_i \right) \right]^{(1-\delta_i)}
\]  

and that the log likelihood of the sample observations is

\[
\log L = - \sum_{i=1}^{n} \delta_i \log(a + bZ_i) - \sum_{i=1}^{n} \delta_i t_i (a + bZ_i)^{-1} \sum_{i=1}^{n} \left( 1 - \delta_i \right) T_1 (a + bZ_i)^{-1}
\]  

where

\[
\delta_i = \begin{cases} 1 & \text{if } T_i \text{ is Uncensored} \\ 0 & \text{if } T_i \text{ is Censored} \end{cases}
\]
The maximum likelihood estimators $a$ and $b$ are obtained as solution to
\[
\frac{\delta \log L}{\delta a} = - \sum_{i=1}^{n} \delta_i (\hat{a} + \hat{b} z_1)_{-1} + \sum_{i=1}^{n} \delta_i t_i \frac{z_1 (\hat{a} + \hat{b} z_1)^{-2}}{1} + \sum_{i=1}^{n} (1-\delta_i) T_i (\hat{a} + \hat{b} z_1) = 0 \tag{3.3.20}
\]
and
\[
\frac{\delta \log L}{\delta b} = - \sum_{i=1}^{n} \delta_i z_1 (\hat{a} + \hat{b} z_1) + \sum_{i=1}^{n} \delta_i t_i z_1 (\hat{a} + \hat{b} z_1)^{-2} + \sum_{i=1}^{n} (1-\delta_i) T_i z_1 (\hat{a} + \hat{b} z_1) = 0 \tag{3.3.21}
\]
The solutions of (3.3.20) and (3.3.21) are analogous to the solutions of (3.3.4) and (3.3.5). Initial estimates $a_0$ and $b_0$ are obtained as the simple regression estimates as in (3.3.5) and (3.3.6), however if $t_i > T_1$. $T_1$ replaces $t_i$. The estimates $a_{k+1}$ and $b_{k+1}$ of $(k+1)$th interaction are obtained by solving the two equations with appropriate modifications to account for censoring.
The asymptotic variance - covariance matrix for $(a, b)$ is obtained from (3.3.10), but the expected values of the second partial derivations are obtained with censoring.

Thus we have
\[
A_0 = \frac{\delta^2}{\delta a^2} \log L = \sum_{i=1}^{n} \delta_i (a+b z_1)^{-2} - 2 \delta_i t_i (a+b z_1)^{-3} - 2 \sum_{i=1}^{n} (1-\delta_i) T_i (a+b z_1)^{-3} \tag{3.3.22}
\]
\[
A_1 = \frac{\delta^2}{\delta a \delta b} \log L = \sum_{i=1}^{n} \delta_i z_1 (a+b z_1)^{-2} - 2 \delta_i t_i z_1 (a+b z_1)^{-3} - 2 \sum_{i=1}^{n} (1-\delta_i) T_i z_1 (a+b z_1)^{-3} \tag{3.3.23}
\]
\[
A_2 = \frac{\delta^2}{\delta b^2} \log L = \sum_{i=1}^{n} \delta_i z_1^2 (a+b z_1)^{-2} - 2 \delta_i t_i z_1^2 (a+b z_1)^{-3} - 2 \sum_{i=1}^{n} (1-\delta_i) T_i z_1^2 (a+b z_1)^{-3} \tag{3.3.24}
\]
For the $i$-th patient the probability of failure before $T_1$ is 

$$(1 - \exp(-\lambda_1 T_1)), \text{ where } \lambda_1 = (a+bZ_1)^{-1}$$

It then follows that

$$E(A_j) = \sum_{i=1}^{n} (1 - \exp(-\lambda_1 T_1) \lambda_1^2 Z_i^j) - 2\sum_{i=1}^{n} \int_{0}^{T_1} \lambda_1 \exp(-\lambda_1 t_1) \lambda_1^3 Z_i^j \, dt_1 - 2\sum_{i=1}^{n} \exp(-\lambda_1 T_1) T_1 \lambda_1^3 Z_i^j, \quad j = 0, 1, 2, \ldots \quad [3.3.25]$$

Since

$$T_1 \int_{0}^{T_1} \exp(-\lambda_1 t_1) t_1 \, dt_1 = \lambda_1^{-2} (1 - \exp(-\lambda_1 T_1)) - \lambda_1 T_1 (\exp(-\lambda_1 T_1)) \quad [3.3.26]$$

We then find

$$E(A_j) = \sum_{i=1}^{n} Z_i^j \lambda_1^2 (1 - \exp(-\lambda_1 T_1)) \quad [3.3.27]$$

Analogous to (3.3.11) it follows that

$$\text{Var}(\hat{a}) = -\frac{E(A_2)}{E(A_2) E(A_0)^2 - [E(A_1)]^2}$$

$$\text{Var}(\hat{b}) = -\frac{E(A_0)}{E(A_2) E(A_0)^2 - [E(A_1)]^2}$$

and

$$\text{Cov}(\hat{a}, \hat{b}) = -\frac{E(A_1)}{E(A_2) E(A_0)^2 - [E(A_1)]^2}$$

If patients enter the clinical trial uniformly during the time interval $(0,T)$, the density function for $T_1$ is

$$g(T_1) = \begin{cases} 1/T & 0 < T_1 < T \\ 0 & \text{elsewhere} \end{cases} \quad [3.3.31]$$

Defining $E^* (A_j)$ as $E_{T_1} E_{T_1} (A_j / T_1)$, we find that

$$E^* (A_j) = -\sum_{i=1}^{n} Z_i^j \lambda_1^2 \left[1 - (\lambda_1 T_1)^{-1} \right] (1 - \exp(-\lambda_1 T_1)) \quad [3.3.32]$$

The formulas for $\text{Var}(\hat{a})$, $\text{Var}(\hat{b})$ and $\text{Cov}(\hat{a}, \hat{b})$ given by (3.3.28) through (3.3.30) change accordingly with $E^* (A_j)$, $j = 0, 1, 2$. 

63
Application: We consider the paraplegia study ambulant times (See Chapter II) and lower limb scores as covariate in the RAD treatment group. For these data, the maximum likelihood estimates are \( \hat{a} = 87.3 \) and \( \hat{b} = -14.7 \). Their corresponding standard errors which are found using (3.3.22) through (3.3.24) directly (not taking expectations) are \( S_{a} = 21.3 \) and \( S_{b} = 4.9 \). These results throw lot of insight into the interpretation of the data. This data is further considered in the subsequent chapters.

3.4 GENERAL PARAMETRIC GROWTH MODELS FOR ASSESSMENT OF SURVIVABILITY

In this section we discuss the distribution of the number of survivors at the end of a clinical trial (which is binomial) and the effects of modifications in the clinical trial on new groups of patients with respect to increasing proportion of survivors.

To investigate these effects, we assume that the clinical trial is carried out in independent stages where in the clinical procedure is modified between successive stages with the aim of increasing the proportion of survivors from stage to stage. To ensure independent stages, each stage of the clinical trial contains a new group of patients.

To this end, suppose a clinical trial is conducted in \( k \) stages such that at the \( i \)-th stage, \( n_i \) patients enter the study each with a probability \( p_i \) of surviving the \( i \)-th stage. At each stage a new group of patients is entered, thus stages are assumed to be independent as well as identical in their time periods.

64
After the 1-th stage, the probability that there are $x_1$ survivors is

$$\Pr\{ X=x_1 \} = \left( \begin{array}{c} n_1 \\ x_1 \end{array} \right) P_1^{x_1} (1 - P_1)^{n_1-x_1} \quad 1 = 1,2,\ldots k.$$  

We discuss in some detail the general parametric form that is assumed by $P_1$ for measuring survivability from stage to stage with some specific examples of the general model. We hypothesize a parametric form for $P_1$ subject to the course of random fluctuations. We assume that $P_1$ is a function of at most two parameters and discuss three specific functions of $P_1$. Model 1 is referred as

$$P_1 = P_\infty - \alpha G(1) \quad [3.4.1]$$

where $P_\infty$, $0<P_\infty<1$, is the ultimate theoretical proportion of survivors achievable, $\alpha > 0$ is a second parameter that quantifies the amount of growth between stages 1 and k and $G(1)$ is known decreasing function of 1. The model for which $G(1)=1/1$ is known as a hyperbolic growth model. Model 2 is given by

$$P_1 = 1 - \alpha_1 \exp(-\alpha_2 l) \quad [3.4.2]$$

where $0 < \alpha_1 < \exp(\alpha_2)$, $\alpha_2 > 0$ are the parameters of the model. This model is termed an exponential growth model. In some applications estimators of $P_\infty$ in Model 1 and $\alpha_1$, $\exp(\alpha_2)$ in Model 2 violate their physical constraints. The fair straightforward remedy for this is Model 3 defined by the equation,

$$P_1 = \{ 1 + \exp \left[ - (\alpha_1+\alpha_2 l) \right] \}^{-1} \quad [3.4.3]$$

where $\alpha_1$ and $\alpha_2$ are the parameters of the model. This model is a special case of the more general class of logistic models.
Generally, \( P_i = P_i(\alpha_1, \alpha_2) \) is a function of two parameters \( \alpha_1 \) and \( \alpha_2 \) and the stage number of the trial is \( i \) where \( i = 1, 2, \ldots k \). Then given that a clinical trial has been conducted in \( k \) stages with \( x_1 \) survivors of \( n_1 \) patients in the first stage, \( x_2 \) survivors of \( n_2 \) patients in the second stage, \( \ldots \), \( x_k \) survivors of \( n_k \) patients in the \( k \)-th stage, the problem is to estimate the parameters \( \alpha_1 \) and \( \alpha_2 \).

Two methods commonly used to estimate \( \alpha_1 \) and \( \alpha_2 \) are the least squares and maximum likelihood. The maximum likelihood estimators are generally preferred to the least squares estimators because of their desirability of the large sample properties. On the other hand, the least squares estimators are often obtainable in closed forms and are a good first approximation to the maximum likelihood estimators.

**Least squares:** Let us define \( \psi(\alpha_1, \alpha_2) \) as

\[
\psi(\alpha_1, \alpha_2) = \sum_{1}^{n} \left( \frac{x_i}{n_i} - P_i(\alpha_1, \alpha_2) \right)^2
\]

[3.4.4]

The least squares estimators \( \hat{\alpha}_1 \) and \( \hat{\alpha}_2 \) are the values of the parameters \( \alpha_1 \) and \( \alpha_2 \) respectively that simultaneously minimise \( \psi(\alpha_1, \alpha_2) \).

**Maximum likelihood:** The likelihood function for the all \( k \) stages is given by

\[
L(\alpha_1, \alpha_2) = \prod_{1}^{n} \left( \frac{n_i}{x_i} \right) \left[ P_i(\alpha_1, \alpha_2) \right]^{x_i} \left[ 1 - P_i(\alpha_1, \alpha_2) \right]^{n_i - x_i}
\]

[3.4.5]

The m.l. estimates \( \hat{\alpha}_1 \) and \( \hat{\alpha}_2 \) are the values of the parameters \( \alpha_1 \) and \( \alpha_2 \) respectively that simultaneously minimize \( L(\alpha_1, \alpha_2) \).
Thus for Model 1 the least square estimator $\mathbf{p}^*_{\infty}$ and $\mathbf{\alpha}^*$ are given by

$$
\mathbf{p}^*_{\infty} = \frac{(\Sigma G^2(1)) \left( \Sigma \frac{x_1}{n_1} \right) - (\Sigma G(1)) \left( \Sigma G(1) \frac{x_1}{n_1} \right)}{k \Sigma G^2(1) - (\Sigma G(1))^2}
$$

$$
\mathbf{\alpha}^* = \frac{(\Sigma G(1)) \left( \Sigma \frac{x_1}{n_1} \right) - k \left( \Sigma G(1) \frac{x_1}{n_1} \right)}{k \Sigma G^2(1) - (\Sigma G(1))^2}
$$

The maximum likelihood estimators $\mathbf{p}^*_{\infty}$ and $\mathbf{\alpha}$ are given by the unique solution to

$$
\sum_{1}^{n} \left( \frac{x_1}{(\mathbf{p}^*_{\infty} - \mathbf{\alpha} G(1))} - \sum_{1}^{n} \frac{n_1 - x_1}{(1-\mathbf{p}^*_{\infty} + \mathbf{\alpha} G(1))} \right) = 0
$$

and

$$
- \sum_{1}^{n} \frac{x_1 G(1)}{(\mathbf{p}^*_{\infty} - \mathbf{\alpha} G(1)) - \sum_{1}^{n} \frac{n_1 - x_1}{G(1)/(1-\mathbf{p}^*_{\infty} + \mathbf{\alpha} G(1))}} = 0
$$

Unfortunately $\mathbf{p}^*_{\infty}$ and $\mathbf{\alpha}$ cannot be obtained in closed forms and can be determined using a two dimensional Newton-Raphson method.

Similarly for Model 2 the least square estimators $\mathbf{\alpha}^*_{1}$ and $\mathbf{\alpha}^*_{2}$ are given by

$$
\log \mathbf{\alpha}^*_{1} = 2 \left( k(k-1) \right)^{-1} \left( \Sigma Z_1 - 3 \Sigma 1 Z_1 \right)
$$

and,

$$
\mathbf{\alpha}^*_{2} = 6 \left( k(k-1) \right)^{-1} \left\{ \Sigma Z_1 - 2 \Sigma 1 Z_1 / (k+1) \right\}
$$

where,

$$
Z_1 = \log \left( \frac{(n_1 - x_1)}{(n_1 + 1)} \right)
$$

The maximum likelihood estimators $\mathbf{\alpha}^*_{1}$ and $\mathbf{\alpha}^*_{2}$ are obtained as the unique solution to

$$
- \sum_{1}^{n} \frac{x_1}{\left\{ \exp (\mathbf{\alpha}^*_{2} - \alpha_1) \right\}} + \sum_{1}^{n} \frac{n_1 - x_1}{\alpha_1} = 0
$$

and,

$$
\mathbf{\alpha}^*_{1} \sum_{1}^{n} \frac{1 x_1}{\left\{ \exp (\mathbf{\alpha}^*_{2} - \alpha_1) \right\}} - \sum_{1}^{n} \frac{1 (n_1 - x_1)}{\alpha_1} = 0
$$

Model 3 also termed as the logistic model can be rewritten as

$$
\log \left( \frac{P_1}{(1-P_1)} \right) = \alpha_1 + \alpha_2 i \quad i = 1, \ldots, k
$$
If the plots the stage by stage log odds - that is \( \log(x_i/(n_i-x_i)) \) and these plot roughly as a straight line as a function of 1, there is empirical justification for utilizing Model 3. The parameters \( \alpha_1 \) and \( \alpha_2 \) can be estimated by maximum likelihood estimators and the procedure is straightforward. However to obtain closed form expressions for estimators of \( \alpha_1 \) and \( \alpha_2 \) and to take advantage of regression we consider a method of estimation as described in Gross and Clark (1975). According to the above procedure the least squares estimators \( \hat{\alpha}_1 \) and \( \hat{\alpha}_2 \) are given by

\[
\hat{\alpha}_1 = \left\{ \frac{n}{1} \left( \sum i^2w_1^2 \right) \left( \sum z_1^2w_1 \right) - \left( \sum i^2w_1 \right) \left( \sum z_1w_1 \right)^2 \right\} / D [3.4.15]
\]

and

\[
\hat{\alpha}_2 = \left\{ \frac{n}{1} \left( \sum w_1^2 \right) \left( \sum z_1^2w_1 \right) - \left( \sum i^2w_1 \right) \left( \sum z_1w_1 \right)^2 \right\} / D [3.4.16]
\]

where \( Z_1 = \log(r_1/(1-r_1)) \), \( r_1 = x_1/n_1 \)

and \( w_1 = (n_1r_1(1-r_1))^{1/2} \), \( D = \left( \sum w_1^2 \right) \left( \sum i^2w_1 \right) - \left( \sum i^2w_1 \right)^2 \)

To obtain a \((1-\alpha)\) 100 percent confidence interval for \( P_{k+1}(\hat{\alpha}_1, \hat{\alpha}_2) \) the predicted probability of survival at the \( (k+1) \)-th stage of the clinical trial we must find \( \text{Var} P_{k+1}(\hat{\alpha}_1, \hat{\alpha}_2) \). Since we may find \( \text{Var} P_1(\hat{\alpha}_1, \hat{\alpha}_2) \) at any stage \( i \) of the clinical trial, we do not restrict our procedure to finding only \( \text{Var} P_{k+1}(\hat{\alpha}_1, \hat{\alpha}_2) \). Using Taylor series the approximate variance of \( P_1(\hat{\alpha}_1, \hat{\alpha}_2) \) is

\[
\text{Var} P_1(\hat{\alpha}_1, \hat{\alpha}_2) = \sigma_{\alpha_1}^2 \theta_{11} + \sigma_{\alpha_2}^2 \theta_{21} + 2 \sigma_{\alpha_1} \sigma_{\alpha_2} \theta_{11} \theta_{21} \quad [3.4.17]
\]

where \( \theta_{m1} = \partial P_1(\alpha_1, \alpha_2) / \partial \alpha_m \) \( m = 1, 2 \)

and the values of \( \sigma_{\alpha_1}^2, \sigma_{\alpha_2}^2 \) and \( \sigma_{\alpha_1} \sigma_{\alpha_2} \) are the elements of the matrix.
\[
\begin{pmatrix}
\sigma_{\alpha_1}^2 & \sigma_{\alpha_1} \sigma_{\alpha_2} \\
\sigma_{\alpha_1} \sigma_{\alpha_2} & \sigma_{\alpha_2}^2
\end{pmatrix}
= \begin{pmatrix}
\sigma^{11} & \sigma^{12} \\
\sigma^{12} & \sigma^{22}
\end{pmatrix}^{-1}
\]

where, \( \sigma_{ij} = -E[\delta^2 \log L(\alpha_1, \alpha_2) / \partial \alpha_i \partial \alpha_j] \)

The (1-\( \alpha \)) 100 % lower confidence limit for \( P_{L1}(\hat{\alpha}_1, \hat{\alpha}_2) \) is given by,

\[
P_{L1}(\hat{\alpha}_1, \hat{\alpha}_2) = P_{L1}(\hat{\alpha}_1, \hat{\alpha}_2) - Z_{1-\alpha}(\text{Var} P_{L1}(\hat{\alpha}_1, \hat{\alpha}_2))^{1/2}
\]

where \( Z_{1-\alpha} \) is the upper (1-\( \alpha \))100 % point of the standard normal distribution.

For Model 1

\[
\sigma^{11} = \sum_{i=1}^{n} \left[ n_i \left/ \left( (P_{\alpha} - \alpha G(1)) (1 - P_{\alpha} + \alpha G(1)) \right) \right. \right] \quad [3.4.20]
\]

\[
\sigma^{22} = \sum_{i=1}^{n} \left[ n_i G^2(1) / \left( (P_{\alpha} - \alpha G(1)) (1 - P_{\alpha} + \alpha G(1)) \right) \right] \quad [3.4.21]
\]

\[
\sigma^{12} = \sum_{i=1}^{n} \left[ n_i G(1) / \left( (P_{\alpha} - \alpha G(1)) (1 - P_{\alpha} + \alpha G(1)) \right) \right] \quad [3.4.22]
\]

\[
\text{Var}(P_{L1}(\hat{\alpha})) = \sigma_{P_{\alpha}}^2 + G^2(1) \sigma_{\alpha}^2 - 2G(1) \sigma_{P_{\alpha}, \alpha}
\]

\[
P_{L1}(P_{\alpha}, \alpha) = (\hat{P}_{\alpha} - \alpha G(1)) - Z_{1-\alpha}(\text{Var} P_{L1}(\hat{\alpha}))^{1/2}
\]

For Model 2

\[
\sigma^{11} = \alpha_1^{-1} \sum_{i=1}^{n} \left( n_i / (\exp(\alpha_2,1) - \alpha_1) \right) \quad [3.4.25]
\]

\[
\sigma^{12} = -\sum_{i=1}^{n} \left( n_i / (\exp(\alpha_2,1) - \alpha_1) \right) \quad [3.4.26]
\]

\[
\sigma^{22} = \alpha_1 \sum_{i=1}^{n} \left( n_i^2 / (\exp(\alpha_2,1) - \alpha_1) \right) \quad [3.4.27]
\]

\[
\text{Var} P_{L1}(\hat{\alpha}_1, \hat{\alpha}_2) = \exp(-2\alpha_2,1) \left[ \sigma_{\alpha_1}^2 + \alpha_1 \sigma_{\alpha_2}^2 - 2 \alpha_1 \sigma_{\alpha_1} \sigma_{\alpha_2} \right] \quad [3.4.28]
\]

\[
P_{L1}(\hat{\alpha}_1, \hat{\alpha}_2) = 1 - \exp(-\hat{\alpha}_2,1) - Z_{1-\alpha}(\text{Var} P_{L1}(\hat{\alpha}_1, \hat{\alpha}_2))^{1/2}
\]

For Model 3 an approximate (1-\( \alpha \))100 % lower confidence limit is

given by

\[
P_{L1}(\hat{\alpha}_1, \hat{\alpha}_2) = \left( 1 + \exp \left( - \left( \hat{\alpha}_1 + \hat{\alpha}_2 \right) - Z_{1-\alpha} \left( \sigma_{\alpha_2}^2 + \sigma_{\alpha_1}^2 - 2 \sigma_{\alpha_1} \sigma_{\alpha_2} \right) \right) \right)^{1/2}
\]

[3.4.29]
where,  
\[ a_m^2 = \frac{\sum (1 \cdot w_{1k})}{\Delta} \quad m=0,1,2 \quad \text{and} \]
\[ \Delta = \left( \sum 1^2 \right) \left( \sum w_{1k}^2 \right) - \left( \sum 1 \cdot w_{1k} \right)^2 \]

The above procedure for model 3 is based on a different procedure. See (Gross and Clark, 1975) for obtaining closed form expressions.

AN APPLICATION

In recent years several short-course regimens of 2 to 7 months duration had been conducted at the tuberculosis Research Centre, Madras to assess the efficacy and remission of treatments with rifampicin on the pulmonary tuberculosis disease. Here we consider four stages with new patients selected at each stage. The proportion of patients achieving bacteriological relapse at each stage are given in Table 3.4.1. The maximum likelihood estimates of the parameters for the models are given in Table 3.4.2 along with least squares estimates. The expected probabilities are given Table 3.4.3. From the table we see that all the three models behave similarly and the expected probabilities are approximately equal and agrees with observed values. However Model 2 seems to be more appropriate than the other two.

3.5 ANALYSIS OF SPINAL TUBERCULOSIS SURVIVAL DATA USING PARAMETRIC MODELS

The Madras Study of the Tuberculosis of the Spine is a collaborative study of Indian Council of Medical Research (Tuberculosis Research Centre, Madras) and the British Medical Research Council in co-operation with six orthopaedic departments.
### TABLE 3.4.1
Non-Relapses Under Short-Course Chemotherapy with Rifampicin in Pulmonary Tuberculosis Patients Over 5-Year Follow-up

<table>
<thead>
<tr>
<th>Stage</th>
<th>Duration (months)</th>
<th>Total Pts</th>
<th>Not Relapsed %</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>72</td>
<td>56</td>
<td>78</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>200</td>
<td>161</td>
<td>81</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>316</td>
<td>295</td>
<td>93</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>132</td>
<td>129</td>
<td>98</td>
</tr>
</tbody>
</table>

### TABLE 3.4.2
Maximum Likelihood and Least Squares Estimates of the Parameters for the Models

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>M.L.E</th>
<th>L.S.E</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>$\phi$</td>
<td>0.998</td>
<td>0.993</td>
</tr>
<tr>
<td></td>
<td>$\alpha$</td>
<td>0.209</td>
<td>0.227</td>
</tr>
<tr>
<td>II</td>
<td>$\alpha_1$</td>
<td>0.492</td>
<td>0.497</td>
</tr>
<tr>
<td></td>
<td>$\alpha_2$</td>
<td>0.813</td>
<td>0.836</td>
</tr>
<tr>
<td>III</td>
<td>$\alpha_1$</td>
<td>--</td>
<td>0.649</td>
</tr>
<tr>
<td></td>
<td>$\alpha_2$</td>
<td>--</td>
<td>0.768</td>
</tr>
</tbody>
</table>

### TABLE 3.4.3
Observed Versus Expected Values Under The 3-Models

<table>
<thead>
<tr>
<th>Stage</th>
<th>Obs.Prob.</th>
<th>Expected Probability</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.778</td>
<td>0.789</td>
<td>0.780</td>
<td>0.805</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.805</td>
<td>0.898</td>
<td>0.814</td>
<td>0.868</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.934</td>
<td>0.940</td>
<td>0.931</td>
<td>0.929</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.977</td>
<td>0.975</td>
<td>0.970</td>
<td>0.978</td>
<td></td>
</tr>
</tbody>
</table>
of the hospitals in Madras. The aim of the study is to assess the efficacy of short course chemotherapy when given alone or combined with radical surgery in the treatment of tuberculosis of the spinal chord without paraplegia.

Patients with active spinal tuberculosis involving the bodies of the thoracic and lumbar vertebral were allocated at random to one of the following three treatments.

(a) RAD6 : Chemotherapy with isoniazid (6mg/kg) plus rifampicin (10-15mg/kg) daily for 6 months plus radical surgery.
(b) AMB6 : As in (a) but without surgery.
(c) AMB9 : As in (b) but chemotherapy for 9 months.

In all 303 patients have been admitted (99 RAD6; 101 AMB6; 103 AMB9). For further details of the study the reader is referred to ICMR/BMRC (1989).

In this section we model survival time (response time) as a function of patients' covariates and surgery status and compare the results obtained using various parametric models. The effects of surgery on response is examined by comparing the hazard functions separately. The objective of the present analysis is to explore the use of parametric models for the hazard function. Recent developments in theory and computation (Aitkin and Clayton, 1980, Laird and Oliver, 1981) make it easy to fit Weibull, extreme value and piecewise exponential models with covariates by using a Poisson representation for the likelihood. All the computations for this analysis were carried out using macros developed in GLIM. For the lognormal distribution a normal regression model is fitted.
to the mean of the log survival time, using the EM algorithm (Dampster et al., 1977, Venkatesan et al., 1988b, 1990d) to accommodate the censored observations.

3.5.1. Notations and methods of analysis

Associated with each patient is a date of admission $T_1$ and a date of favourable response $T_2$ and for the RAD6 patients there is also a date of surgery $T_3$ with $T_1 \leq T_3 \leq T_2$. The waiting time for the RAD6 patients is defined as $W = T_3 - T_1$ and their post surgery response as $Y = T_2 - T_3$. For AMB6 and AMB9 series patients we define response time to be $Y = T_2 - T_1$. Associated with each patient there is a vector of covariates $Z$, the age on admission, level of the disease, vertebrae involved etc.

We employ fully parametric modeling using exponential, Weibull, extreme value, lognormal and piecewise exponential distribution. The piecewise exponential approach may be considered as a simple approximation to Cox's (1972) semi-parametric model (see Chapter V) and also a parametric model valid under weaker assumptions than those of fully parametric models. The hazard function for all the models considered except for the lognormal, may be written as

$$h(t) = h_0(t) \exp (\beta'Z). \quad [3.5.1]$$

where $h_0(t)$ has the form

$$h_0(t) = \begin{cases} 
1 & \text{for exponential} \\
\alpha t^{\alpha - 1} & \text{for Weibull, } \alpha > 0 \\
\alpha \exp(\alpha t) & \text{for extreme value, } \alpha > 0 \\
\exp(\pi_j) & \text{for } c_j \leq t \leq c_{j+1}
\end{cases} \quad [3.5.2]$$

72
where the $c_j$ are a set of defined cut points on time axis, $Z$ is a vector of covariates including a vector of all 1's that corresponds to the intercept $\beta_0$. The extreme value distribution has monotone increasing hazard, while the Weibull distribution has monotone increasing hazard for $\alpha > 1$ and monotone decreasing hazard for $0 < \alpha < 1$. The piecewise exponential allows for an arbitrary shape, within the restriction of constant hazard in intervals.

For the lognormal, we assume the log of the survival time $t$ is normal, with mean $\mu = \gamma'Z$ and variance $\sigma^2$. The lognormal has hazard $0$ at $t=0$, rises to a maximum, then decreases monotonically as $t \to \infty$. Because the covariates do not multiply the hazard proportionately the regression coefficients $\gamma$ are not directly comparable with $\beta$'s in the previously mentioned models. The comparison of the Weibull and lognormal models is most conveniently achieved through the median survival time. For the lognormal distribution the median is $\exp(\mu) = \exp(\gamma'Z)$ while for the Weibull distribution the median is $(\log 2)^{1/\alpha} \exp(-\beta'Z/\alpha)$. The median comparison is more appropriate than the mean. Since the different tail behaviour of the lognormal and Weibull distribution may give very large differences in means for the same medians (the means are $\exp(\gamma'Z + \sigma^2/2)$ and $\Gamma(\alpha^{-1}+1)\exp(-\beta'Z/\alpha)$ respectively). For further details and references, the reader is referred to Aitkin and Francis (1983). These relations suggest that the Weibull co-efficient divided by $-\alpha$ can be compared with the corresponding lognormal co-efficient.
For the piecewise exponential, the time axis is divided into k intervals \([C_j, C_{j+1}]\) \(j=0, \ldots, k-1\) with \(C_0 = 0\) and \(C_k = \infty\). Then for the jth interval we define an indicator variable \(Y_{ij}\) for the jth patient which takes the value 1 if the patient experiences the event of interest (e.g. death) in the jth interval, and the value 0 otherwise. Then the Poisson representation in Aitkin and Clayton (1980) applies, time \(Y_{ij}\) being \((nk)\) independent Poisson variables with means \(\mu_{ij}\) given by

\[
\log \mu_{ij} = \log E_{ij} + \pi_j + \beta' Z_i
\]

where \(E_{ij}\), the exposure in the jth interval is defined by

\[
E_{ij} = \begin{cases} 
C_{j+1} - C_j & \text{if the patient is alive at } C_{j+1} \\
t_j - C_j & \text{if the patient dies or censored at } t_j \text{ in } \{C_j, C_{j+1}\} \\
0 & \text{if the patient responded before } C_j
\end{cases}
\]

In the last case, \(\log E_{ij}\) is taken to be zero.

If the covariates \(Z\) are categorical, the analysis reduces to that of fitting a log-linear model to a multiway contingency table, and the standard iterative scaling algorithm may be used (David and Oliver 1981, Holford 1980). The contingency table analysis is particularly useful because it allows the inclusion of non-proportional hazards by extending the lognormal model to include time interval by covariate interactions. Time dependent covariates may also be included in such analysis.

To choose between parametric representation we use significance tests based on differences in minimised deviances (\(-2\log likelihood\)). The term deviance is used since its minimised
value is used to construct hypothesis tests and to compare set of related models. The difference between the exponential and piecewise exponential deviances is asymptotically distributed as \( \chi^2_{k-1} \) when the true hazard is constant. Likewise since the Weibull distribution can be obtained via power transformation of the exponential, the difference in their deviances is asymptotically distributed \( \chi^2_1 \) when the piecewise exponential deviances should be approximately \( \chi^2_{k-2} \) if the time hazard is Weibull even though they are not members of the same parametric family.

3.5.2 Results

Table 3.5.1 gives parameter estimates for the model with main effects of all covariates: age, level, kyposis, limb movements, inactivation, pretreatment, surgery, sex etc., using five different shapes for the underlying hazard: extreme value, Weibull, exponential, piecewise exponential and log normal. The fittings were done using GLIM package. The bottom line gives the minimized deviances for the five models. By comparing the deviances it is clear that neither extreme value, lognormal nor exponential model provides good fit of the data. The Weibull and piecewise exponential are roughly equivalent in terms of deviances \( (\chi^2 = 7.1) \); but exponential differs significantly from Weibull \( (\chi^2 = 29.7) \) and piecewise exponential \( (\chi^2 = 42.8) \).

The estimated regression coefficients for the Weibull and piecewise exponential models are quite similar as are their estimated standard errors. By comparing the results for different equations using only only the main effects we find age, ptb,
<table>
<thead>
<tr>
<th>Para.</th>
<th>Lognormal</th>
<th>Exponential</th>
<th>Piecewise Exp.</th>
<th>Extrevalue</th>
<th>Weibull</th>
</tr>
</thead>
<tbody>
<tr>
<td>INT</td>
<td>5.467(0.550)</td>
<td>-9.764(0.906)</td>
<td>-5.694(0.916)</td>
<td>-5.716(0.902)</td>
<td>-5.918(0.907)</td>
</tr>
<tr>
<td>BGP</td>
<td>-0.046(0.042)</td>
<td>0.075(0.063)</td>
<td>0.086(0.067)</td>
<td>0.095(0.071)</td>
<td>0.076(0.069)</td>
</tr>
<tr>
<td>KYP</td>
<td>-0.111(0.063)</td>
<td>0.235(0.112)</td>
<td>0.132(0.122)</td>
<td>0.219(0.132)</td>
<td>0.226(0.117)</td>
</tr>
<tr>
<td>LEV</td>
<td>0.062(0.083)</td>
<td>-0.017(0.091)</td>
<td>-0.086(0.081)</td>
<td>-0.091(0.081)</td>
<td>-0.164(0.089)</td>
</tr>
<tr>
<td>AGE</td>
<td>-0.042(0.039)</td>
<td>0.216(0.072)</td>
<td>0.107(0.061)</td>
<td>0.129(0.069)</td>
<td>0.117(0.071)</td>
</tr>
<tr>
<td>PTB</td>
<td>-0.112(0.061)</td>
<td>0.301(0.088)</td>
<td>0.216(0.096)</td>
<td>0.401(0.102)</td>
<td>0.441(0.093)</td>
</tr>
<tr>
<td>SUR</td>
<td>-0.069(0.022)</td>
<td>0.140(0.032)</td>
<td>0.148(0.031)</td>
<td>0.134(0.029)</td>
<td>0.167(0.032)</td>
</tr>
<tr>
<td>INS</td>
<td>-0.151(0.097)</td>
<td>0.198(0.117)</td>
<td>0.191(0.101)</td>
<td>0.099(0.113)</td>
<td>0.081(0.109)</td>
</tr>
<tr>
<td>LIM</td>
<td>-0.066(0.073)</td>
<td>0.117(0.034)</td>
<td>0.139(0.041)</td>
<td>0.114(0.034)</td>
<td>0.168(0.033)</td>
</tr>
<tr>
<td>PRX</td>
<td>-0.101(0.041)</td>
<td>0.096(0.056)</td>
<td>0.127(0.050)</td>
<td>0.087(0.063)</td>
<td>0.061(0.057)</td>
</tr>
<tr>
<td>SEX</td>
<td>0.082(0.031)</td>
<td>-0.076(0.033)</td>
<td>-0.098(0.025)</td>
<td>-0.109(0.034)</td>
<td>-0.112(0.035)</td>
</tr>
</tbody>
</table>

DEVIAN. | 457.1 | 481.3 | 431.6 | 781.9 | 438.7 |

INT-Intercept  BGP-Blood Group  KYP-Kyposis  LEV-Disease Level  PTB-Pulmonary Tb.
SUR-Surgery  INS-Inactivation  LIM-Lim Movements  PRX-Previous Rx

\[
\begin{align*}
\text{BLOCK(2)} &= -1.967(0.751) \\
\text{BLOCK(3)} &= -1.489(0.566) \\
\text{BLOCK(4)} &= -1.355(0.676) \\
\text{BLOCK(5)} &= -2.139(0.641) \\
\text{BLOCK(6)} &= -2.119(0.534)
\end{align*}
\]
surgery, limb movements and pretreatment contribute significantly. We categorized the time axis into six intervals defined by
\[ t < 60; \ 60 \leq t < 120; \ 120 \leq t < 180; \ 180 \leq t < 240; \ 240 \leq t < 300; \ t \geq 300 \text{ days}. \]
The coefficients corresponding to the blocks are given in the bottom of the table. The log normal distribution fitted to response time eventhough there is no real indications from the piecewise exponential of a non-monotone hazard. Using deviances from the lognormal fit to test different models leads to the same qualitative conclusion about the effects of covariates.

In conclusion, it seems worthwhile to see that regression analysis with censored data have the same potential problems of ordinary regression analysis in addition to others arising as a result of the censoring. Our analysis as pointed out the relative insensitivity of our results to the model used to specify the effects of covariates.