INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6” x 9” black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
A covariate model in finite mixture survival distributions

Soegiarso, Restuti Widayati, Ph.D.
Case Western Reserve University (Health Sciences), 1992
A COVARIATE MODEL
IN FINITE MIXTURE SURVIVAL DISTRIBUTIONS

by

RESTUTI WIDAYATI SOEGIARSO

Submitted in partial fulfillment of the requirements
for the Degree of Doctor of Philosophy

Thesis Advisor: Dr. Nahida H. Gordon

Department of Epidemiology and Biostatistics
CASE WESTERN RESERVE UNIVERSITY
MAY 1992
CASE WESTERN RESERVE UNIVERSITY
GRADUATE STUDIES

We hereby approve the thesis of

RESINTI WIDAYATI SOEIARSO

candidate for the PH.D.
degree.*

Signed: \underline{\text{Paula H. Gooden}}

(chairman)

\underline{\text{Paul K. Jones}}

Date 3/20/92

*We also certify that written approval has been obtained for any proprietary material contained therein.
I grant to Case Western Reserve University the right to use this work, irrespective of any copyright, for the University's own purposes without cost to the University or to its students, agents and employees. I further agree that the University may reproduce and provide single copies of the work, in any format other than in or from microforms, to the public for the cost of reproduction.

[Signature]
A COVARIATE MODEL
IN FINITE MIXTURE SURVIVAL DISTRIBUTIONS

Abstract
by
RESTUTI WIDAYATI SOEGIARSO

Within a clinical trial setting, observations often come from different classes or groups that are related to the disease being studied. These classes may be different diagnoses or prognoses that may affect survival. Furthermore, some observations may have unknown group memberships, or new observations come along, and it is of interest to classify these observations into one of the groups based on their known survival times and covariates. This research addresses the problem of using finite mixture distributions with covariates to model such heterogeneous data.

The survival distribution of these observations is considered as a mixture of two component exponential distributions, each representing a group. The hazard function of each component distribution is expressed as a log-linear function of the covariates. Maximum likelihood estimates of the covariate coefficients and the parameters of the component distributions are obtained. Five procedures that use the survival times as well as the covariates are constructed for classifying new observations into one of the component distributions. The classification procedures are compared with the neighborhood non-parametric procedure that uses the covariates alone.
The methodology is illustrated using data from a prospective clinical trial in breast cancer based at Case Western Reserve University. These patients are considered to be a mixture of those with one to three positive nodes, and those with four or more positive nodes. Two covariates, estrogen receptor status and tumor diameter, are used in the model. The survival distributions from the proposed model seem to fit the actuarial distribution quite well. The proposed classification procedures also work quite well, and in fact they produce smaller error rates than the neighborhood procedure.

To study the performances of the classification procedures when different sizes of correlations between the grouping variable and the covariates exist, the methodology is applied to simulated clinical trial data. The results from the simulation study suggest that when small or moderate correlation exists, the classification procedures that use the survival times and the covariates are more accurate than the neighborhood method. Furthermore, when the correlation is large, the proposed classification procedures yield better or comparable results to the neighborhood method.
TO MY PARENTS,

AND

IN MEMORY OF MY BELOVED BROTHER,

SANTO NUGROHO SOEGIARSO
ACKNOWLEDGMENTS

I would like to express my sincere thanks to my advisor, Dr. Nahida H. Gordon, for her supervision and expertise in this research. I am very grateful for her encouragement, guidance and support throughout my graduate studies at Case Western Reserve University. Her professional and personal attention are greatly appreciated. I would also like to thank her for being such a great mentor and teacher.

I would like to thank my committee members, Dr. Sara M. Debanne and Dr. Paul K. Jones from the Epidemiology and Biostatistics Department; and Dr. Joseph P. Crowe, Jr. from the Department of Surgery for their inputs and constructive suggestions throughout the different stages of my dissertation. I am grateful for the positive responses they have always shown towards my research.

Many thanks also go to Dr. Joseph P. Crowe, Jr. and Dr. Nahida H. Gordon for providing me with the Breast Cancer Research Project data base to apply the methodology, and the Epidemiology and Biostatistics Department, especially Dr. Harold B. Houser, for their financial support during my Ph.D. program at Case Western Reserve University.

I would like to acknowledge my parents, Ir. Soegiarso Padmopranoto and Mrs. Sri Koeswati Soegiarso in Jakarta, Indonesia. I am extremely grateful that they have provided me with such a good home and an intellectual environment to grow up in. I am also very grateful for their encouragement and support for all these years. I would never be where I am right now without their love and guidance.
Many thanks go to everyone in my family for their moral support: my sisters, Dina E. Agustiono, Ir. Ina Soegiarso and Peni H. Hariadi, M.B.A; my brothers in-law, Dr. Emil Agustiono and Drs. Heru Hariadi; my niece, Wulan E. Agustiono; and my nephews: Aji E. Agustiono, Arif E. Agustiono and Timur H. Hariadi.

I would like to express special thanks to my fiancée, Dr. Moula Cherikh, who has been giving me encouragement and support. Many thanks also go to all my friends in Cleveland, especially Edith Berger, Jane Brumberg and Gladys Stefanek, who have made my study and stay in Cleveland a more enjoyable one.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>DEDICATION</td>
<td>iv</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>v</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>vii</td>
</tr>
<tr>
<td>CHAPTER 1. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>CHAPTER 2. LITERATURE REVIEW</td>
<td>4</td>
</tr>
<tr>
<td>2.1. Finite Mixtures</td>
<td>4</td>
</tr>
<tr>
<td>2.1.1. Introduction and Examples</td>
<td>4</td>
</tr>
<tr>
<td>2.1.2. Definition of Finite Mixtures</td>
<td>8</td>
</tr>
<tr>
<td>2.1.3. Identifiability</td>
<td>9</td>
</tr>
<tr>
<td>2.1.4. Sampling</td>
<td>10</td>
</tr>
<tr>
<td>2.1.5. Methods of Estimation</td>
<td>12</td>
</tr>
<tr>
<td>2.1.5.1. Definition and Description</td>
<td>12</td>
</tr>
<tr>
<td>2.1.5.2. Comparison of Methods and Discussion</td>
<td>21</td>
</tr>
<tr>
<td>2.2. Covariates</td>
<td>24</td>
</tr>
<tr>
<td>2.2.1. Introduction</td>
<td>24</td>
</tr>
<tr>
<td>2.2.2. Statistical Methods for Multivariate Techniques</td>
<td>26</td>
</tr>
<tr>
<td>2.3. Censoring Mechanism</td>
<td>33</td>
</tr>
<tr>
<td>CHAPTER 3. THE MIXTURE MODEL</td>
<td>36</td>
</tr>
<tr>
<td>3.1. Introduction and Objectives</td>
<td>36</td>
</tr>
<tr>
<td>3.2. The Mixture Model</td>
<td>38</td>
</tr>
<tr>
<td>3.3. Comparisons of Models</td>
<td>43</td>
</tr>
<tr>
<td>3.3. Classification Procedures</td>
<td>42</td>
</tr>
</tbody>
</table>
CHAPTER 1
INTRODUCTION

Within a clinical trial setting, each observation in the study population can be one of different types or classes. Measurements on these observations are known to belong to one of a set of classes, but individual class memberships may be unavailable or indistinguishable. So instead of having a single distribution, these measurements have a distribution which is a mixture of two or more distributions. The distributions that are mixed generally belong to the same family but with different parameter values. In such an event we are dealing with heterogeneous populations having a finite mixture distribution.

As an example, a population of patients with a certain disease might be composed of two component populations: one component will die of the disease and the other component will die of something else. In this case the overall distribution of survival times of these patients has a finite mixture distribution, specifically a mixture of two distributions. Usually, the data are available only for the overall mixture distribution, and not for each component, therefore it is of interest to estimate the parameters of the survival functions of each population component as well as the proportion of each component. This estimation gives the ability to understand the behavior of the population distribution function in terms of its components. Since the results of clinical trials are frequently complete only after a long period of follow-up, and it is generally of interest to estimate the proportion of patients cured or who will die from the disease after some treatment, this ability is especially useful in a long-term follow-up trial.

However, traditionally finite mixture methods do not take into account all
other variables that might influence the outcome of a disease. These other variables are called the covariates or prognostic variables. Covariates or prognostic variables are usually collected prior to or during clinical trials. As an example, in breast cancer clinical trials these covariates can be age at diagnosis, tumor diameter, number of nodes, estrogen receptors, and menopausal status at the time of surgery. Covariates can be very important in predicting the outcome of a disease. In most cases, outcome is defined as the disease-free time, or recurrence free-time, or survival time from the disease, or total survival time. Covariate studies have been used widely in the area of survival data analysis. The knowledge of how the covariates affect the outcome of a disease is useful in making a prediction on the patient’s outcome based upon the patient’s covariates.

Traditionally, finite mixture methods are applied to either survival or covariate data. This research attempts to incorporate covariates into finite mixture survival methods, so that a new patient can be classified into one of the component populations based on that patient’s covariates. The model proposed in this research is different from the classical discriminant analysis, because the proposed model takes into account both censored and non-censored survival times and covariates.

Traditionally, classical discriminant methods do not include the survival data, but are only applied to covariate data. Covariates can be incorporated into mixture survival distributions by expressing the hazard function of each component population in the training sample as a function of the covariates. The parameters of the mixture distribution in the training sample are expressed as functions of the covariate coefficients. Therefore a new observation with certain covariate values can be classified into one of the component populations through a probability rule.

In this research, finite mixture densities are interpreted as densities
associated with a statistical population which is a mixture of component populations or classes each with its own density function. Mixtures of this kind have received considerable attention in the area of survival analysis. The covariates considered in this research are of the fixed type, i.e., their values do not change with time.
CHAPTER 2
LITERATURE REVIEW

2.1. Finite Mixtures

2.1.1. Introduction and Examples

The term "mixtures" or "mixture models" or "mixed distributions" has been used rather loosely in the area of survival analysis by some investigators. Many use this term to describe the population being studied, even though they do not use the theory of finite mixtures. For example, Boag (1949), Berkson and Gage (1952), Cutler and Axtell (1963), Haybittle (1965), and Pocock. Gore and Kerr (1982) have considered groups of cancer patients as composed of two components, those who will die of other causes with no evidence of disease relapse and those who will die of their disease, but they did not use distribution functions from the same family. Boag (1949) used the log-normal distribution to model the survival function of patients who died of cancer and used the information on disease status of patients who were still alive to estimate the proportion cured and the parameters of the log-normal distribution. Berkson and Gage (1952) used census data and actuarial methods to estimate the death rate of the 'cured' component and assumed that the cancer component followed an exponential distribution. Cutler and Axtell (1963) and Pocock. Gore, and Kerr (1982) used actuarial methods to estimate the survival functions.

Farewell (1982) also used the term "mixture model" even though he did not use the theory of finite mixtures. He tried to relate time to death data to toxicant levels and other stresses on laboratory animals. The time to death among animals that died early due to experimental stresses was modelled as a Weibull distribution. The distribution of time to death as a function of the stresses was represented as a logistic
model. The mixture model of the probability of death that he proposed was a product of the time to death among animals that will die early and the distribution of time to death as a function of the stresses.

The following examples illustrate the use of the theory of finite mixtures that will be applied in this proposed work. The interest in finite mixtures began at the end of the nineteenth century when Karl Pearson (1894) estimated the five parameters in a mixture of two normal distributions. Pearson's method of moments have been extended to more general mixtures of normal densities and to mixtures of other continuous densities by several authors. Pollard (1934) obtained moment estimates for a mixture of three univariate normal densities. Cooper (1967), Day (1969), and John (1970) studied moment estimates for mixtures of multivariate normal densities. They simplified the assumptions about the mixtures under consideration to reduce the complexity of the moment estimation problem.

Rider (1961) derived moment estimates for the means in a mixture of two exponential densities under the assumption that the mixing proportions are known. Later in 1962, he derived moment estimates for mixtures of Weibull distributions. John (1970) studied moment estimates for mixtures of two gamma distributions. Fractional moments in mixtures of two exponential distributions were explored by Tallis and Light (1968).

Everitt and Hand (1981) reviewed the literature in finite mixtures and indicated the practical details of fitting such distributions to sample data. They also discussed the properties and the estimation methods for mixtures of normal, exponential, and other continuous distributions.

Medgyessi (1961) analyzed absorption spectra and protein separation by electrophoresis in terms of normal mixtures. Cassie (1954) and Bhattacharya (1967)
studied the length distribution of a certain type of fish and found that splitting the observations into age categories with each category contributing a normal component was useful in getting the overall mixture. Gregor (1969) applied a mixture of normal distribution to the measurement data of the content of DNA in the nuclei of liver cells of rats. Examples of normal mixtures that are used to investigate the robustness of certain statistical techniques when the data are not normally distributed were given by Subrahmaniam, Subrahmaniam and Messeri (1975), and Hyrenius (1950). Clark et al. (1968) studied disease distributions to answer the question whether there is more than one type of disease in hypertension. Specifically, they studied whether a sample of blood pressure data can be separated into two normal distributions.

Failure data, where the observations are the times to failure of a sample of items, is also another area where mixture distributions are important. Davis (1952) studied the failure that can occur for more than one reason and approximated the failure distribution for each reason by a simple density function such as the negative exponential. The overall distribution was then a mixture of some negative exponential distributions. Mendenhall and Hader (1958) fitted exponential components to the failure distribution of communication transmitter receivers of a single commercial airline. The data was censored at a predetermined test termination time. Kao (1959) fitted Weibull components to the distribution of electronic valves.

Failure can also be interpreted as an event. The distribution of time to discharge of nerve cells was studied by Thomas (1966) as an exponential mixture. Ashton (1971) used a mixture of a gamma distribution to model the frequency distribution of time gaps in road traffic.

Inzenman and Sommer (1988) discussed the parametric and nonparametric perspectives of the mixture problem and its special application to identifying the
components of "philatelic mixtures". They described the term "philatelic mixtures" as any situation in which a particular stamp issue is known to have been printed on a mixture of paper types with possible differences in paper thickness.

Gordon (1989, 1990) applied the theory of finite mixtures to estimate the proportion cured in population of women with breast cancer. She used a mixture of two Gompertz distributions to model the overall death distribution, one component represented death due to breast cancer and the second component represented death due to other causes. The Gompertz distribution has never been used in finite mixtures before.

Titterington, Smith and Makov (1985) covered the aspects of finite mixtures in detail, such as definitions and basic concepts, and applications of finite mixtures. They discussed the mathematical aspects of finite mixtures and the estimation methods of the parameters, such as the graphical method, method of moments, maximum likelihood, bayesian methods and minimum distance methods.

The properties of the finite mixtures where the parameters are estimated using different methods have also been investigated by several authors. Hill (1963) investigated the Fisher information for estimating the proportion \( p \) in a mixture of two exponential and two normal densities. He concluded that extremely large and impractical sample sizes are needed to obtain even moderate precision in estimating \( p \) unless the mixed distributions are very well separated. Hosmer (1973) presented the results of a Monte Carlo study of the maximum likelihood estimates for selected cases of mixtures of two univariate normal distributions with \( \sigma_1 \neq \sigma_2 \), \( |\mu_1 - \mu_2| < 3 \min(\sigma_1, \sigma_2) \), and with sample sizes, \( N \leq 300 \). He found that larger samples are needed to derive stable estimates. He also concluded that, although the maximum likelihood method is the best available, it may not be highly satisfactory for even moderate
sample sizes, and thus should be used with extreme caution.

Dick and Bowden (1973) applied the maximum likelihood method to estimate the five parameters in the mixture of two normal distributions using the Newton-Raphson iterative method when independent sample information is available from one of the populations. They then used the Monte Carlo simulation to obtain the sample variances of the estimates as well as the estimated asymptotic variances. They found that when the number of observations is small and the means are not well separated, the sample variance of the estimates can be as much as three times greater than the estimated asymptotic variance.

Gordon (1989, 1990) studied the feasibility of finite mixture theory in a population of cancer patients for estimating the proportion cured after treatment. The Newton-Raphson method was used to solve the maximum likelihood equation for estimating the parameters. The population of cancer patients was considered as a mixture of two Gompertz components. She found that a moderate sample size that is typical of cooperative trials yields acceptable confidence ellipsoids for the proportion cured and other parameters. She also concluded that data which are as much as 50% censored yield acceptable results, and moreover, data which are as much as 70% censored can yield satisfactory results for some parameters.

2.1.2. Definition of Finite Mixture Distributions

Finite mixture densities can be interpreted as densities associated with a statistical population which is a mixture of m component populations or subpopulations with associated component densities \( \{ f_i \}_{i=1}^m \) and mixing proportions \( \{ p_i \}_{i=1}^m \). Titterington, Smith and Makov (1985) defined a random variable \( X \) as having a finite mixture distribution if its probability density function is of the form:
\( f(x) = p_1 f_1(x) + p_2 f_2(x) + \ldots + p_m f_m(x) \quad (x \in X) \)

where \( X \) is the range of \( X \), \( p_i > 0, \; i=1,\ldots,m \), \( \sum_{i=1}^{m} p_i = 1 \) and \( f_i(.) \geq 0 \). The parameters \( p_1,\ldots,p_m \) are called the mixing weights or proportions and \( f_1(.),\ldots,f_m(.) \) the component densities of the mixture. In many situations \( f_1(.),\ldots,f_m(.) \) will belong to a parametric family of distributions and so the mixture distribution will have the more explicit representation:

\[ f(x) = p_1 f_1(x|\theta_1) + \ldots + p_m f_m(x|\theta_m) \quad (2.1.2.1) \]

where \( \theta_i \) denotes the parameter occurring in \( f_i(.) \).

Let \( \Phi \) denote the complete collection of all distinct parameters in the mixture model. Thus \( \Phi = \{(p_i, \theta_i); \; i=1,\ldots,m\} \). The problem here is estimating \( \Phi \).

2.1.3. Identifiability

The concept of identifiability is important in the theory of finite mixtures. Without the assumption of identifiability, estimation procedures can not be well defined. In general, identifiability is the existence of a unique parameterization for any one of the class of distributions being considered.

Everitt and Hand (1981) gave the following definition of identifiability. A class \( D \) of mixtures is said to be identifiable if and only if for all \( f(x) \in D \) the equality of two representations

\[ \sum_{i=1}^{c} p_i f_i(x; \theta_i) = \sum_{j=1}^{c'} p'_j f'_j(x; \theta'_j) \]

implies that \( c = c' \) and for all \( i \) there exists some \( j \) such that \( p_i = p'_j \) and \( \theta_i = \theta'_j \).
As an example, let \( U(x; a,b) \) denote the uniform density on \((a,b)\). Then, \( U(x; 0,1)=p \ U(x; 0,p) + (1-p) \ U(x; p,1) \) is not identifiable, because for any \( p \) between 0 and 1, the mixture always yields the same result.

Identifiability is important in estimation, because multiple solutions can make the estimation of parameters much more difficult. Furthermore, this is a more serious problem when we need to classify future observations into one of the components distributions. This problem is known as the non-identifiability of a class of mixtures. Before attempting to estimate the parameters in finite mixture distributions one has to make sure that the class of distributions being considered is identifiable. Teicher (1961, 1963) first studied this problem.

Teicher (1961) established that a mixture of Poisson distributions is identifiable, while binomial and uniform families do not produce identifiable mixtures. Teicher (1963) also showed that finite mixtures of one-dimensional normal distributions and gamma distributions are identifiable. Yakowitz and Spragins (1968) proved that \( n \)-products of exponential distribution, multi-dimensional normal distribution, union of exponential and normal families, family of one-dimensional Cauchy distributions, and the non-degenerate members of the family of one-dimensional negative binomial distributions are identifiable. Members of the exponential family of distributions are also identifiable.

2.1.4. Sampling

A sample observation on the mixture is labeled if its component population is known with certainty, and is unlabeled otherwise. Estimating \( \Phi = \{(p_i, \theta_i): i=1,\ldots,m\} \) is done through a sample in which some or all of the observations are unlabeled.
There are varieties of samples that might arise in mixture problems. Redner and Walker (1984) reviewed this problem and introduced four types of samples, denoted as Type 1, 2, 3, and 4. A sample of Type 1, $S_1$, is composed of a random sample of $N$ unlabeled observations $\{x_k\}_{k=1,\ldots,N}$ from the mixture distribution with density $f(x|\Phi)$. A practical example would be a sample of survival times of patients whose disease category is unknown. So the samples are completely unlabeled, and the proportion as well as the parameters of each component are unknown. In this case, whether or not the labeling is caused by censoring is unknown.

Suppose that $J_1, \ldots, J_m$ are arbitrary nonnegative integers and for $i=1,\ldots,m$, $\{y_{ik}\}_{k=1,\ldots,J_i}$ is an independent sample of observations from the $i$th component population, i.e., a set of $J_i$ observations on independent, identically distributed random variables with density $f_i(x|\theta_i)$. Then $S_2 = \bigcup_{i=1}^{m} \{y_{ik}\}_{k=1,\ldots,J_i}$ is a sample of Type 2. As an example, a sample of survival times from different disease categories is available. In this case, the samples are completely labeled and there is no censoring.

In a Type 3 sample, one has an independent sample of $K$ unlabeled observations drawn from the mixture. These observations are subsequently labeled. And for $i=1,\ldots,m$, a set $\{z_{ik}\}_{k=1,\ldots,K_i}$ of them is associated with the $i$th component population with $K = \sum_{i=1}^{m} K_i$. Then $S_3 = \bigcup_{i=1}^{m} \{z_{ik}\}$ is a sample of Type 3. For example, one has a sample of survival times from different disease categories, where subsequently the disease category becomes known.

In the last type of sample, one has an independent sample of $M$ unlabeled observations drawn from the mixture. The observations in the sample which fall in some set $E \subseteq \mathbb{R}^n$ are subsequently labeled so that for $i=1,\ldots,m$ a set $\{w_{ik}\}_{k=1,\ldots,M_i}$ of them is associated with the $i$th component population, while a set $\{w_{0k}\}_{k=1,\ldots,M_0}$ remains unlabeled. Then $S_4 = \bigcup_{i=0}^{m} \{w_{ik}\}_{k=1,\ldots,M_i}$ is a sample of Type 4. An example
of Type 4 sample would be a sample of survival times from different categories, from which some of the disease categories are known and some are not known due to censoring.

A totally unlabeled sample of Type 1 is considered in almost all of the literature on mixtures. In Type 3 the numbers $K_i$ contain information about the mixing proportions, while the numbers $J_i$ in Type 2 do not. A sample of Type 4 is encountered most of the time in practice, especially in clinical settings which involve censored data. The numbers $M_i$ in a sample of Type 4 contain information about the mixing proportions, and also about the parameters of the component densities.

Tan and Chang (1972) used a sample of the form of $S_1 \cup S_2$ to explain a genetic variation problem. Hosmer (1973) considered samples of the forms $S_1$, $S_1 \cup S_2$, and $S_1 \cup S_3$. A sample of the form $S_1 \cup S_2$ where $m=2$ and $J_2=0$ was considered by Dick and Bowden (1973). Hosmer and Dick (1978) evaluated the Fisher information matrix for a variety of samples of Type 1, 2, 3, and their unions. A sample of Type 4 was used by Mendenhall and Hader (1958) for a mixture of exponential distributions. Gordon (1989, 1990) used a sample of Type 4 to describe a mixture population of treated cancer patients. Her study population was composed of $M$ observations which represent patients with cancer, in which $M_1$ of them will die of their cancer, $M_2$ of them will die of other causes, and $M_0$ of them remains unlabeled because of censoring. In this proposed research, the model of the finite mixtures will be adapted from the Type 4 sample.

2.1.5. Methods of Estimation

2.1.5.1. Definition and Description

Estimation methods for mixtures and properties of these methods have been
considered widely in the literature, ranging from Pearson’s method of moments (1894) to graphical methods. In the following section the maximum likelihood method and other methods will be discussed.

A. Maximum Likelihood Method

This is the most widely used method for estimating the parameters in finite mixture densities. The likelihood function of a sample of observations is the probability density function of the random sample evaluated at the observations at hand. It is usually more convenient to work with the log-likelihood function rather than with the likelihood function. Let \( \theta \) denote the collection of all distinct parameters in the component densities. The log-likelihood functions \( LL_1, LL_2, LL_3 \) and \( LL_4 \) of samples \( S_1, S_2, S_3 \) and \( S_4 \) of Types 1, 2, 3, and 4 respectively are found in Redner and Walker (1984):

\[
LL_1(\Phi) = \sum_{k=1}^{N} \log f(x_k|\Phi)
\]

\[
LL_2(\Phi) = \sum_{i=1}^{m} \sum_{k=1}^{J_i} \log f_i(y_{ik}|\theta_i)
\]

\[
LL_3(\Phi) = \sum_{i=1}^{m} \sum_{k=1}^{K_i} \log [p_i f_i(z_{ik}|\theta_i)] + \log \frac{K_i!}{K_i^1 \ldots K_i^m}
\]

\[
LL_4(\Phi) = \sum_{k=1}^{M_0} \log f(w_{0k}|\Phi) + \sum_{i=1}^{m} \sum_{k=1}^{M_i} \log [p_i f_i(w_{ik}|\theta_i)] + \log \frac{M_i!}{M_0^1 \ldots M_m^i}
\]

The likelihood functions of Type 1, Type 2, Type 3 and Type 4 samples will be denoted by \( L_1, L_2, L_3, \) and \( L_4 \) respectively, which are simply the exponential of the corresponding log-likelihood. The maximum likelihood estimate \( \Phi^* \) is any choice of \( \Phi \).
in $\Omega$, where $\Omega$ is the set of allowable values for $\Phi$, at which the log-likelihood function of $S$, denoted by $LL(\Phi)$ attains its largest value in $\Omega$. In other words, a maximum likelihood estimate associated with a sample of observations is a choice of parameter which maximizes the probability density function of the sample, called the likelihood function. For many parameter estimation problems this maximization can be tackled by differentiating the logarithm of the likelihood function with respect to the parameters, $\Phi$, and setting the derivatives equal to zero, to give the likelihood or normal equations

$$\frac{dLL(\Phi)}{d\Phi} = 0.$$  

These are then solved for $\Phi$, and the second-order derivatives are evaluated to verify that it is indeed a maximum which has been achieved and not some other stationary point.

For mixture distribution problems, however, the likelihood equations are almost certain to be nonlinear and very difficult to solve analytically. Consequently, some iterative procedures are needed for an approximate solution. There are several general iterative procedures which are suitable for finding an approximate solution to the likelihood equations. The maximum likelihood estimates can be solved by numerical methods such as: a) EM algorithm, b) Newton-Raphson method, and c) method of scoring. In the following section these three numerical methods are reviewed.

a. **EM Algorithm**

Dempster, Laird and Rubin (1977) presented the EM algorithm as an iterative computation procedure of maximum likelihood estimates when the observations can be viewed as incomplete data. This method is called the EM algorithm since each
iteration of the algorithm consists of an Expectation step followed by a Maximization step.

The first step of the iteration involves estimating the membership probabilities of each observation for each component. The second step of the iteration is equivalent to m separate estimation problems with each observation contributing to the m components with a weight given by the estimated membership probability. These two steps are then repeated iteratively.

The general EM algorithm works as follows. Suppose one has to find \( \Phi = \tilde{\Phi} \) to maximize the likelihood \( L(\Phi) = f(x|\Phi) \), where \( x \) is a set of 'incomplete data'. In the context of mixtures, 'incomplete data' refer to the unlabeled samples, and the 'complete data' refer to the labeled samples. Let \( y \) denote a typical 'complete' version of \( x \) and let \( Y(x) \) denote the set of all possible such \( y \). Let the likelihood from \( y \) be denoted by \( g(y|\Phi) \). The EM algorithm generates, from some initial approximation, \( \Phi^{(t)} \), a sequence \( \{ \Phi^{(s)} \} \) of estimates. Each iteration consists of the following two steps:

E step: Evaluate \( E[\log g(y|\Phi)| x, \Phi^{(s)}] = Q(\Phi, \Phi^{(s)}) \).

and

M step: Find \( \Phi = \Phi^{(s+1)} \) to maximize \( Q(\Phi, \Phi^{(s)}) \).

It has been found in most instances that the EM algorithm has the advantage of having reliable global convergence, low cost per iteration, economy of storage and ease of programming. Unfortunately its convergence can be extremely slow in simple problems which are often encountered. The lack of speed of the EM method has been commented on by many authors. Everitt and Hand (1981) quoted convergence rates for a sample of 200 observations drawn from a four components Poisson mixture. From one set of starting values the EM algorithm takes 192 iterations, from another.
Furthermore, the two sets of final values are different, suggesting the existence of multiple maxima.

Redner and Walker (1984) applied the EM algorithm to normal mixtures. They discovered that although convergence typically takes very many iterations, it does not take long to get 'high-up' on the likelihood surface. Usually 95 percent of the change in the log-likelihood can be achieved in the first five iterations. They suggested that a composite algorithm, in which a few EM iterations are followed by one or two further iterations using the Newton-Raphson method or the method of scoring, would be worth investigating.

b. Newton-Raphson Method

Let $LL(\Phi)$ denote the log-likelihood of interest. Then the iterative step for the Newton-Raphson algorithm is defined by:

$$\Phi^{(s+1)} = \Phi^{(s)} - \left[ D^2 \ LL(\Phi^{(s)}) \right]^{-1} D \ LL(\Phi^{(s)}), \quad s = 0, 1, \ldots$$

where $D$ is the first differentiation with respect to $\Phi$ and $D^2$ is the second differentiation with respect to $\Phi$.

The Newton-Raphson method can be difficult to use, particularly in view of the matrix inversion and that there is no guarantee of monotonicity. However, the Newton-Raphson method converges quite fast unless the separation between the components is poor. Furthermore, if the asymptotic theory holds, the iterations for the Newton-Raphson method incorporates approximations to the covariance matrix of $\Phi$, the maximum likelihood estimator of $\Phi$ which is $\left[ -D^2 \ LL(\Phi) \right]^{-1}$. 
c. **Method of Scoring**

Let $LL(\phi)$ be the log-likelihood of interest. Then the iterative step for the method of scoring algorithm is defined by:

$$\phi^{(s+1)} = \phi^{(s)} - \left[ I(\phi^{(s)}) \right]^{-1} D L(\phi^{(s)}), \quad s = 0, 1, \ldots$$

where $D$ is the first differentiation with respect to $\phi$ and $I(\phi)$ is the Fisher information matrix which is $E \left[ D L(\phi) D L(\phi)' \right]$.

As with the Newton-Raphson method, the method of scoring can be difficult to use in view of the matrix inversion, and there is no guarantee of monotonicity. The method of scoring involves an integration in evaluating $I(\phi)$, which can be a difficult numerical problem.

B. **Brief Description of Other Methods**

The method of moments was used as early as the nineteenth century when Pearson (1894) attempted to estimate the five parameters of a two-component univariate normal mixture. This method was one of the most popular ways of estimating the parameters in a mixture distribution before computers became available.

The method of moments consists of equating some set of sample moments to their expected values or population moments to obtain a system of equations for the parameters in the mixture density. This system of equations is generally non-linear. Consider a two-component mixture distribution $f(x) = p f_1(x) + (1-p) f_2(x)$, with $r$ parameters. Then by equating the observed moments given by

$$V_s = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^s \quad s = 1, \ldots, r$$

to the theoretical first $r$ moments given by
\[ v_s = \int (x - \mu)^s f(x) \, dx \quad s = 1, \ldots, r. \]

where \( \mu = E(x) \), a system of \( r \) non-linear simultaneous equations can be obtained, which must be solved to give estimates of the \( r \) parameters in the mixture distribution. Explicit solution to this method may not be easy or even possible. In general numerical methods are needed to obtain solutions.

In addition to the estimation methods cited above, a variety of other methods have been proposed for estimating parameters in mixture densities, ranging from a Bayesian method to a graphical method.

The Bayesian method is based on prior information about the sample data. Let \( L(\Phi) \) denote the likelihood for \( \Phi \) given sample data \( x = (x_1, \ldots, x_n) \). Bayes theorem uses beliefs about \( \Phi \) prior to observing \( x \) which is expressed as a density \( f(\Phi) \). This prior information is then updated into beliefs about \( \Phi \) posterior to observing \( x \) denoted by \( p(\Phi|x) \) and given by

\[
p(\Phi|x) = \frac{L(\Phi) f(\Phi)}{\int L(\Phi) f(\Phi) \, d\Phi}.
\]

(2.1.5.1)

In the context of finite mixture models, \( L(\Phi) \) could take any of the forms \( L_1(\Phi) \), \( L_2(\Phi) \), \( L_3(\Phi) \), \( L_4(\Phi) \) described in the previous section, with necessary modifications if some of the component parameters are known.

Another method that has been used in finite mixture estimation is the minimum distance method based on distribution functions. This comprises a variety of procedures for estimating a probability density function or a distribution function. Suppose \( F(.|\Phi) \) is the distribution function of interest with the unknown parameters \( \Phi \). Let \( F_n(.) \) denote the empirical distribution function obtained from a sample of \( n \) independent observations. \( x \), and \( \delta(G, F) \) be a measure of distance between two
distribution functions, \( F \) and \( G \). Then the minimum distance estimator \( \hat{\Phi} \) for \( \Phi \) is a value of \( \Phi \) which minimizes \( \delta(F_n(\cdot), F(\cdot|\hat{\Phi})) \). If this distance is denoted by \( \delta(\hat{\Phi}) \) for convenience, a necessary condition for \( \hat{\Phi} \) to be a relative minimum of \( \delta(\Phi) \) is to satisfy

\[
D_{\Phi} \delta(\hat{\Phi}) = 0.
\]

where \( D \) is the first differentiation with respect to \( \Phi \). Thus this approach chooses parameter estimates so that the estimated mixture distribution could be as close as possible to the data.

The minimum distance method can also be based on the transforms of the distribution function, such as the moment generating function. Quandt and Ramsey (1978) derived the moment generating function method to estimate the parameters in a mixture of two univariate normal distributions. The procedure selects the parameters which minimize the sum of squares of differences between the theoretical moment generating function denoted by

\[
M(\beta) = E(\exp \beta X),
\]

which depends on the parameters \( \Phi \), and the sample moment generating function

\[
\frac{1}{n} \sum_{i=1}^{n} \exp (\beta X_i)
\]

i.e. by minimizing the expression

\[
e(\Phi) = \sum_{j=1}^{m} \left[ M(\beta_j) - \sum_{i=1}^{n} \exp(\beta_j x_i)/n \right]^2
\]

for chosen values of \( \beta_1, \beta_2, \ldots, \beta_m \).

In the case of a mixture of two univariate normal distributions derived by Quandt and Ramsey (1978), the theoretical moment generating function \( M(\beta) \) will take
the form

$$E \exp(\beta x) = p \exp(\mu_1 \beta + \sigma_1^2 \beta^2/2) + (1-p) \exp(\mu_2 \beta + \sigma_2^2 \beta^2/2).$$

For any given value of $\beta$, say $\beta_j$, the moment generating function may be estimated by the sample moment

$$\hat{E} \exp(\beta_j x) = \frac{1}{n} \sum_{j=1}^{n} \exp(\beta_j x_j).$$

The last estimation method that will be discussed in this section is the graphical method. The graphical procedure using probability paper is a simple way of resolving finite mixture densities and is usually done to provide crude estimates of the underlying unknown parameters. This method was popular before computers became widely available. Cox (1966) suggested the following approach for mixtures of two components of known distribution.

In order for this approach to work well, the two components have to be well separated, i.e. greater number of points in the lower tail of the observed distribution will come from one of the components, say $f_1$, and the majority of points in the upper tail will come from $f_2$. In such a case plots on probability paper of (i) the number of observations less than $x$ divided by $np$, and (ii) the number of observations greater than $x$ divided by $n(1-p)$, should produce straight lines if $x$ is either small enough or large enough. This approach can be easily applied to distributions for which simple methods of probability plotting are available such as normal, Gamma, and Weibull mixtures. Most of the published work is concerned with univariate data and mixtures of normal or log-normal densities.
2.1.5.2. Comparisons of Methods and Discussion

The properties and performance of the maximum likelihood method and other methods have been investigated by several authors. In the present section the comparisons of the methods discussed previously and the choice of method for this proposed research will be presented.

Redner and Walker (1984) and Everitt and Hand (1980) stated that the maximum likelihood method and the method of moments have been the most popular methods for estimating the parameters of finite mixture distributions. However, with the availability of powerful computers and sophisticated numerical methods developed during the 1960’s, the maximum likelihood method became preferred to the method of moments for mixture density estimation problems, and in fact in all areas of statistical estimation. The maximum likelihood method has desirable statistical properties, that is, under very general conditions the estimators are consistent and they are asymptotically normally distributed.

When using any algorithm (such as the EM algorithm, Newton-Raphson, or the method of scoring) to find the maximum likelihood estimator of Φ, there is the practical difficulty that the algorithm might stop at a local optimum rather than the global maximum. To avoid this, one might run the algorithm several times with different starting points for Φ each time and then hopefully one may be able to choose the global maximum among these several local optima.

The method of moments is simpler to use, though explicit solution through this method may not also be easy or even possible. Numerical methods are generally needed to solve the non-linear system of equations. Several authors have compared the efficiency of the method of moments relative to the maximum likelihood method under different circumstances.
Kendall and Stuart (1963) found that the sampling variance of a moment depends on the population moment of twice the order, that is the sample variance becomes very large for higher moments even when the sample size is very large. Therefore the method of moments will not be practical for a large number of parameters, or multivariate problems.

Cohen (1967), Day (1967), and Tan and Chang (1970) have examined the method of moments. Their simulation studies have shown that in the univariate case the moment estimates should only be used with large sample sizes. For a sample size of 1000 they are of sufficient accuracy to provide only an initial guess for the maximum likelihood estimates. Tan and Chang (1970) studied the joint asymptotic relative efficiency comparisons and showed that the maximum likelihood estimates are much more efficient than the moment estimates, especially when \( |\mu_2 - \mu_1| \leq 2\min(\sigma_1, \sigma_2) \). Day (1969) showed that the moment estimates are essentially useless for multidimensional mixtures. Everitt and Hand (1981) stated that it seems fair to say that usually the moment estimators are inefficient.

Titterington, Smith and Makov (1985) suggested that the implementation of the Bayesian method in finite mixtures will not be straightforward at all, unless \( \Phi \) consists of one or two unknown parameters. When \( \Phi \) consists of three or more unknown parameters, the problem of performing efficient numerical integration in several dimensions to obtain the denominator of equation (2.1.5.1) is encountered.

Wayne et al. (1984) compared the minimum distance method based on the distribution function and maximum likelihood method for estimating the parameters in a mixture of two normal distributions. Their simulation results indicated that the maximum likelihood method is superior to the minimum distance method when component distributions are actually normal, but the minimum distance method will
provide better estimates than maximum likelihood under symmetric departures from component normality.

Quandt and Ramsey (1978) did a simulation study and showed that the moment generating function method applied to a mixture of two univariate normal distributions compares favorably with the method of moments. Further simulation results by Hosmer (1978) indicated that this method also outperformed maximum likelihood in many situations. But simulation work by Kumar, Nicklin and Paulson (1979) showed that the differences in the various studies are produced because Quandt and Ramsey (1978) and Hosmer (1978) used the true parameter values as initial estimates in the algorithm used to minimize the sum of squares of differences between the theoretical and the sample moment generating functions. Everitt and Hand (1981) confirmed the result of Kumar et al. (1979). They suggested that the moment generating function procedure is far more sensitive to choice of starting values than the maximum likelihood method when applied to mixtures of more than two univariate normal distributions. They also concluded that though the moment generating function method is an interesting approach to estimate the parameters in a mixture, in general it is less satisfactory than the maximum likelihood method.

The graphical techniques give accurate parameter estimates only if the components are well separated or a large sample size is available. However, they may be useful for the initial examination of data since these techniques can be used without a prior knowledge of the number of components in the mixture. Furthermore, they may provide crude estimates of the underlying parameters.

Based on the above discussion, the maximum likelihood method seems to be superior than the other estimation methods. Furthermore, this method has been widely used in finite mixture estimation as well as in classical statistical estimation and
hypothesis testing. Therefore the choice of estimation method in this proposed research is the maximum likelihood.

2.2. Covariates

2.2.1. Introduction

In medical practice, prognosis of a patient plays an important role. Prognosis is the prediction of the future of an individual patient with respect to duration, course, and outcome of a disease. Before prognosis of a patient can be predicted, a medical history as well as pathologic, clinical and laboratory data are often needed. All of those variables that contain patient characteristics are called covariates, prognostic factors, concomitant variables, or independent variables.

In a clinical trials setting, covariates or prognostic variables are usually collected prior to or during the trial. These covariates are usually useful in predicting or affecting the outcome of the trial and are also another way of handling heterogeneity in a population. The outcome variable is some measure of how an individual patient with the disease does. Examples of outcome variables are response to therapy (yes or no), survival time, time to disease recurrence, and relapse-free survival.

A covariate may be either numerical or non-numerical. Numerical covariates may be discrete, such as the number of lymph nodes, or continuous, such as blood pressure. Continuous covariates can be transformed into discrete covariates by grouping patients into subcategories, for example two estrogen receptor (ER) subgroups: ER~ if estrogen receptors are $< 3.0$ femtomoles/mg, and ER+ if estrogen receptors are $\geq 3$ femtomoles/mg. Non-numerical covariates may be ordered, such as severity of disease may be primary, local, or metastatic, or unordered, such as race. They can also be dichotomous, such as the liver is or is not enlarged. Usually, the covariates collected
include some of each type.

Some covariates may change over time. If a covariate changes over time, it is called a "time-dependent" covariate, otherwise it is called a "fixed" covariate. A time dependent covariate may itself be an outcome variable, and so interpretation may not be easy. The analytical methods for handling this type of covariate are more complicated than for fixed covariates. A fixed covariate has a single value for each patient and is determined at the outset of the study. Fixed covariates are all known at the same time prior to observing the outcome, so they cannot be easily confused with the outcome variable. In this proposed research all covariates will be assumed to be of the fixed type.

One purpose of conducting a covariate study is to predict, with some accuracy, the outcome of the disease. It may also provide some insight into the disease process and provide directions for further study. Covariate analysis attempts to relate covariates to the outcome variable and each other. Covariate studies have been done quite extensively in the area of survival analysis.

Usually the number of covariates collected in these studies is very large. Therefore the data have to be screened to remove some covariates from further consideration, leaving a relatively smaller number of covariates that can be investigated more intensely. If some of the covariates are highly correlated, then one of these covariates is likely to be as good a predictor as all of them. Correlation coefficients between covariates can be computed to detect covariates that are significantly correlated. In dropping any highly correlated covariates, information from previous or other studies has to be incorporated. Some covariates may also be dropped because of the quality of the data, or because there are too many missing values. George (1988) suggested that one rule of thumb is to eliminate any covariate with at least 20% of
values missing.

A univariate approach is another method for screening the covariates. It can be performed on each covariate, one at a time, to help to identify which covariates are closely related to the outcome variable. In one approach, the data are grouped according into categories defined by the breakpoints of each covariate, and then the estimated survival functions or hazard functions of patients in the different categories are obtained and compared. A life table analysis may be performed separately for each group when there are many patients in each subgroup. However, this could be very tedious with a large number of covariates. The Kaplan-Meier product limit method is often used to estimate the survival function for each group and each covariate. Survival times of subgroups can then be compared by the standard non-parametric methods such as the Gehan's generalized Wilcoxon test, Cox-Mantel test, log-rank test, or the Kruskal-Wallis test. For continuous covariates, Cox's regression model can also be used for each covariate one at a time.

There are some disadvantages in using these univariate approaches. A large number of patients will be required in each subgroup, and there is no indication of the relative importance of the covariates. Furthermore, one may also miss the correlation that might exist between covariates. Therefore, the univariate technique can only be used to provide a preliminary idea of which covariate has prognostic importance. The multivariate techniques have to be performed to investigate simultaneously the effects of the covariates on the outcome variable. These techniques will be discussed in the following statistical methods section.

2.2.2. Statistical Methods for Multivariate Techniques

The multiple regression method is a conventional technique for investigating
the simultaneous effects of the covariates on the outcome variable. Let \( z_1, z_2, \ldots, z_u \) be \( u \) possible covariates. For the \( k \)th patient, observed values of the covariates are \( z_{1k}, z_{2k}, \ldots, z_{uk} \). In the multiple regression approach the outcome variable, \( x_k \), is the dependent variable. The problem is identifying a relationship of \( x_k \) or a function of \( x_k \), say \( h(x_k) \) and \( (z_{1k}, z_{2k}, \ldots, z_{uk}) \) that can be expressed in a regression function \( x_k = g_1(z_{1k}, z_{2k}, \ldots, z_{uk}) \), or \( h(x_k) = g_2(z_{1k}, z_{2k}, \ldots, z_{uk}) \). Two examples of \( g_1 \) and \( g_2 \) are:

\[
x_k = \beta_1 z_{1k} + \beta_2 z_{2k} + \ldots + \beta_u z_{uk} \quad \text{and} \quad h(x_k) = \exp(\beta_1 z_{1k} + \beta_2 z_{2k} + \ldots + \beta_u z_{uk}).
\]

The traditional least-square method based on the normal distribution can not always be adapted for the survival data with censoring. Lehman (1953) stated that regression models proposed for survival distributions generally involve the assumption of proportional hazards. Cox (1972) introduced a general nonparametric model for the analysis of survival data with and without censoring. In recent years, the parametric regression methods have also been proposed by some investigators. These regression models are based on the assumption that the underlying survival distribution is exponential. The log-linear exponential regression model was introduced by Glasser (1967). The linear exponential model was considered by Feigl and Zelen (1965) and extended to include censored data by Zippin and Armitage (1966). Bryar et al. (1974) also introduced the linear exponential model.

**Cox's Regression Model**

Let the outcome variable \( t \) be the survival time, and \( h_k(t) \) be the hazard function of the \( k \)th patient. When the survival times are continuously distributed and
there are no ties among them, the hazard function is

\[ h_k(t) = h_0(t) \exp(\beta_1 z_{1k} + \beta_2 z_{2k} + \ldots + \beta_u z_{uk}) \]  \hspace{1cm} (2.2.2.1)

where \( h_0(t) \) is the hazard function of the underlying survival distribution (arbitrary) when all the \( z \) variables are ignored, or all \( z \)'s are equal to zero, and \( \beta \)'s are the regression coefficients. Cox's model assumes that the hazard of the study group is proportional to that of the underlying distribution \( h_0(t) \).

Solving the above equation as a regression problem can be done by dividing both sides of the equation by \( h_0(t) \) to obtain

\[ \log \frac{h_k(t)}{h_0(t)} = \beta_1 z_{1k} + \beta_2 z_{2k} + \ldots + \beta_u z_{uk} = \sum_{j=1}^{u} \beta_j z_{jk} \] \hspace{1cm} (2.2.2.2)

The left hand side of equation (2.2.2.2) is a function of the hazard for the \( k \)th patient and the right hand side is a linear combination of the covariates with the coefficients. The \( z \)'s can be continuous covariates, time-dependent covariates, or indicator variables. Letting \( y_k = \log [h_k(t)/h_0(t)] \), equation (2.2.2.2) becomes

\[ y_k = \beta_1 z_{1k} + \beta_2 z_{2k} + \ldots + \beta_u z_{uk} \] \hspace{1cm} (2.2.2.3)

which is simply a standard multiple regression equation with the covariates as independent variables and a function of the hazard as dependent variable.

The stepwise regression method can be applied to solve the equation and to rank the covariates in order of relative importance to the outcome variable. Cox's regression model can also be used to define a prognostic index or hazard ratio, defined as \( \log [h_k(t)/h_0(t)] \) for each patient. This ratio is useful to compare two treatment groups or prognoses between patients.

The independent variables in equation (2.2.2.2) can also be standardized
about the mean to obtain the following equation

\[
\log \frac{h_k(t)}{h_0(t)} = \beta_1(z_{1k} - \bar{z}_1) + \ldots + \beta_u(z_{uk} - \bar{z}_u) \tag{2.2.2.4}
\]

where \( \bar{z}_u \) is the average of the \( u \) th covariate for all patients. Then \( h_0(t) \) is the hazard function when all variables are at their average values. The hazard ratio is then the ratio of risk of failure for a patient with a given set of values \( z_{1k}, z_{2k}, \ldots, z_{uk} \) to that for an average patient who has an average value for every covariate. Model (2.2.2.4) is more realistic and easier to interpret than model (2.2.2).

Cox suggested a maximum likelihood procedure to estimate the coefficients \( \beta_1, \ldots, \beta_u \). The likelihood function is based on a conditional probability of failure. Let \( t_{(1)} < t_{(2)} < \ldots < t_{(r)} \) be the \( r \) exact failure times, and \( R(t_{(s)}) \) be the risk set at time \( t_{(s)} \). \( R(t_{(s)}) \) consists of all individual patients whose survival times are at least \( t_{(s)} \). For the particular failure at time \( t_{(s)} \), conditionally on the risk set \( R(t_{(s)}) \) the probability that the failure is on the individual as observed is

\[
\frac{\exp \left( \sum_{j=1}^{u} \beta_j z_{jk} \right)}{\sum_{i \in R(t_{(s)})} \exp \left( \sum_{j=1}^{u} \beta_j z_{ij} \right)}.
\]

Each failure will contribute a factor and hence the conditional log-likelihood is

\[
LL(\beta) = \sum_{k=1}^{r} \sum_{j=1}^{u} \beta_j z_{jk} - \sum_{k=1}^{r} \sum_{i \in R(t_{(s)})} \exp \left( \sum_{j=1}^{u} \beta_j z_{ij} \right) \tag{2.2.2.5}.
\]

The maximum likelihood estimates of \( \beta \)'s are the estimates that maximize \( LL(\beta) \). They can be obtained by solving simultaneously the equations \( \partial LL(\beta) / \partial \beta_i = 0 \), \( i = 1, \ldots, u \). Some iterative procedures such as the Newton-Raphson method, may be needed to solve the \( u \) equations in (2.2.2.5).
Log-Linear Exponential Model

Suppose there are \( n = n_1 + \ldots + n_m \) individuals in \( m \) groups. Let \( t_{ik} \) be the survival time and \( z_{ik1}, z_{ik2}, \ldots, z_{iku} \) the covariates of the \( k \)th individual in the \( i \)th group, where \( u \) is the number of covariates considered, \( i = 1, \ldots, m \), and \( k = 1, \ldots, n_i \).

This model assumes that the underlying distribution is exponential and the survival function for the \( k \)th individual in the \( i \)th group is

\[
S_{ik}(t) = e^{-\lambda_{ik}t},
\]

where the hazard

\[
\lambda_{ik} = \exp(a_i + \sum_{j=1}^{u} b_jz_{ikj}).
\]

The term \( \exp(a_i) \) represents the underlying hazard of the \( i \)th group when covariates are ignored. The following indicator variables can be used to distinguish censored observations from the uncensored:

\[
\delta_{ik} = \begin{cases} 
1, & \text{if } t_{ik} \text{ is uncensored} \\
0, & \text{if } t_{ik} \text{ is censored.}
\end{cases}
\]

The likelihood function is then

\[
L(\lambda_{ik}) = \prod_{i=1}^{m} \prod_{k=1}^{n_i} (\lambda_{ik})^{\delta_{ik}} e^{-\lambda_{ik}t_{ik}}
\]

and the log-likelihood function of \( a = (a_1, \ldots, a_m) \) and \( b = (b_1, \ldots, b_u) \) is

\[
LL(a,b) = \sum_{i=1}^{m} \sum_{k=1}^{n_i} \delta_{ik} \left( a_i + \sum_{j=1}^{u} b_jz_{ikj} \right) - t_{ik} \exp\left( a_i + \sum_{j=1}^{u} b_jz_{ikj} \right)
\]

\[
= \sum_{i=1}^{m} \left[ a_i \delta_{i1} + \sum_{j=1}^{u} b_j s_{ij} - e^{a_i} \sum_{k=1}^{n_i} t_{ik} \exp\left( \sum_{j=1}^{u} b_jz_{ikj} \right) \right]
\]

where \( s_{ij} = \sum_{k=1}^{n_i} \delta_{ik} z_{ikj} \) is the sum of the \( j \)th covariate corresponding to the uncensored
survival times in the ith group and \( t_i \) is the number of uncensored times in that group.

Maximum likelihood estimates of \( a \) and \( b \) can be obtained by solving the \((k+p)\) equations simultaneously, which can be done by using iterative procedures such as the Newton-Raphson method.

**Linear Exponential Model by Feigl and Zelen**

A. Uncensored Data

Let \( t_1, \ldots, t_N \) be a sample of \( N \) survival times from patients whose individual survival distribution is exponential. Let \( z_{kj} \), \( k=1,\ldots,N, j=1,\ldots,u \) be the observed value of the \( j \)th covariate such that the mean survival time of the \( k \)th patient is a linear function of the \( z_{kj} \)'s, i.e.,

\[
\mu = \frac{1}{\lambda_k} = b_0 + b_1 z_{k1} + \ldots + b_u z_{ku} = \sum_{j=0}^{u} b_j z_{kj}
\]

where \( z_{k0} = 1 \), and \( \lambda_k \) is the hazard of the \( k \)th patient. The term \( b_0 \) represents the underlying hazard, in the sense that \( 1/b_0 \) is the hazard rate. \( \lambda_k \), when covariates are ignored or all \( z_{kj} \)'s are zero. In the regression problem, \( b_0 \) represents the intercept.

The likelihood function of the \( N \) survival times can be written as

\[
L(b) = \prod_{k=1}^{N} f_k(t) = \prod_{k=1}^{N} \lambda_k e^{-\lambda_k t_k}
\]

\[
= \prod_{k=1}^{N} \left( \sum_{j=0}^{u} b_j z_{kj} \right)^{-1} \exp \left[ - \sum_{k=1}^{N} t_k \left( \sum_{j=0}^{u} b_j z_{kj} \right) - 1 \right].
\]

The log-likelihood function is

\[
LL(b) = - \sum_{k=1}^{N} \log \left( \sum_{j=0}^{u} b_j z_{kj} \right) - \sum_{k=1}^{N} t_k \left( \sum_{j=0}^{u} b_j z_{kj} \right) - 1.
\]

The maximum likelihood estimates of \( b = (b_0, b_1, \ldots, b_u) \) can be obtained by solving
simultaneously the \((u+1)\) equations. This can be done by an iterative procedure such as the Newton-Raphson method.

B. Censored Data

Zippin and Armitage (1966) generalized the exponential model proposed by Feigl and Zelen to include censored observations. Suppose \(N\) patients are entered into a study, \(r\) of these die, and \(s = N - r\) are still alive at the end of the study. Let \(t_1, ..., t_r\) be the exact survival times of the \(r\) deaths and \(t_1^+, ..., t_s^+\) be the \(s\) censoring times. Then the likelihood function

\[
L(\lambda) = \prod_{i=1}^{r} \lambda_i e^{-\lambda_i t_i} \prod_{k=1}^{s} e^{-\lambda_k t_k^+}
\]

under the linear exponential model becomes

\[
L(b) = \prod_{i=1}^{r} \left( \sum_{j=1}^{u} b_j z_{ij} \right)^{-1} \exp(-t_i \left( \sum_{j=1}^{u} b_j z_{ij} \right) - 1) \prod_{k=1}^{s} \exp(-t_k^+ \left( \sum_{j=1}^{u} b_j z_{kj} \right) - 1).
\]

The log-likelihood is then

\[
LL(b) = - \sum_{i=1}^{r} \log \left( \sum_{j=1}^{u} b_j z_{ij} \right) - \sum_{i=1}^{r} t_i \left( \sum_{j=1}^{u} b_j z_{ij} \right) - 1 - \sum_{k=1}^{s} t_k^+ \left( \sum_{j=1}^{u} b_j z_{kj} \right) - 1.
\]

The maximum likelihood estimators \(b\) can be obtained by solving simultaneously the \((u+1)\) equations. Again, this can be done by iterative procedures.

Linear Exponential Model by Bryar et al.

Bryar et al. (1974) proposed another exponential model relating censored survival data to covariates for prostate cancer patients in which the individual hazard is linearly related to possible covariates.

Let \(t_1, ..., t_N\) be the survival times of \(N\) individual patients, and \(z_{k0}, z_{k1}, ...,\)
$z_{k0}$ be the $u$ covariates for $k$th individual patient. Supposing the survival distribution is exponential, Bryar et al. suggested that the individual hazard $\lambda_k$ is constant in time and linearly related to the covariates, i.e.,

$$\lambda_k = b_0z_{k0} + b_1z_{k1} + \ldots + b_uz_{ku} = \sum_{j=0}^{u} b_j z_{kj}$$

where $z_{k0}$ is set to one for all patients, $z_{kj}$, is set to 1 if patient $k$ has the covariate $j$ and 0 otherwise, $j \neq 0$, $j=1,\ldots,u$. The intercept $b_0$ represents the underlying hazard rate when all the covariates are ignored.

Suppose that $r$ individuals of the $N$ patients are dead and $s = N - r$ are still alive at the end of the study, then the likelihood function is

$$L = \prod_{i=1}^{r} \lambda_i e^{-\lambda_i t_i} \prod_{k=1}^{s} e^{-\lambda_k t_k^+}$$

$$= \prod_{i=1}^{r} \left( \sum_{j=0}^{u} b_j z_{ij} \right) \exp \left( -\left( \sum_{j=0}^{u} b_j z_{ij} \right) t_i \right) \prod_{k=1}^{s} \exp \left( -\left( \sum_{j=0}^{u} b_j z_{kj} \right) t_k^+ \right).$$

The log-likelihood is

$$LL(b) = \sum_{i=1}^{r} \left[ \log \left( \sum_{j=0}^{u} b_j z_{ij} \right) - \left( \sum_{j=0}^{u} b_j z_{ij} \right) t_i \right] - \sum_{k=1}^{s} \left( \sum_{j=0}^{u} b_j z_{kj} \right)^{-1} t_k^+.$$

The maximum likelihood estimators of $b$ can be obtained by solving simultaneously the $(u+1)$ equations which can be done by iterative procedures.

2.3. Censoring Mechanism

In survival data, the exact lifetimes of some of the individuals may not be observed but are known to exceed certain values. This property is called censoring, and it arises in various ways. Gross and Clark (1975), Kalbfleisch and Prentice (1980), and
Lawless (1982) have provided extensive discussions on censoring types. In this section, the censoring type and mechanism used in this research are discussed.

An observation is said to be right censored at \( L \) if the exact value of the observation is not known but only that it is greater than or equal to \( L \). Similarly, an observation is said to be left censored at \( L \) if it is known only that the observation is less than or equal to \( L \). Right censoring is more common in clinical data than left censoring. In clinical trials, random censoring is also typically used. Random censorship model includes the special case of type I censoring, where the censoring time of each individual is fixed in advance, as well as the case where individuals enter the study at random over time and the analysis is carried out at some prespecified time. Furthermore, the censoring times are usually assumed to be independent of the failure times and their distribution does not involve the parameters of interest. This kind of censoring mechanism is called non-informative. In this research, the lifetime data are assumed to be right censored and the censoring mechanism is non-informative. This situation is typical of breast cancer clinical data which will subsequently be used as an example for the proposed methodology.

Let \( (T_i, L_i) \), \( i=1,\ldots,n \) be independent random variables, where \( T_i \) is the lifetime of the \( i \)th individual with survivor and density functions \( S(t_i) \) and \( f(t_i) \), respectively, with parameter \( \theta_i \); and \( L_i \) is the censoring time of the \( i \)th individual with survivor and density functions \( G(t_i) \) and \( g(t_i) \), respectively, assumed to be independent of the parameter of interest, \( \theta_i \) (i.e. the censoring mechanism is non-informative). Now define \( Y_i = \min(T_i, L_i) \) and

\[
\delta_i = \begin{cases} 
1, & \text{if } T_i \leq L_i \quad \text{(ith individual is not censored)} \\
0, & \text{if } T_i > L_i \quad \text{(ith individual is censored)}.
\end{cases}
\]

The data from observations on \( n \) individuals will consist of the pairs \((y_i, \delta_i)\), \( i=1,\ldots,n \).
where $y_i$ are realizations of $Y_i$. The joint p.d.f. of $(y_i, \delta_i)$ is easily obtained as follows.

$$P(t_i < y_i < t_i + dt, \delta_i = 0) = P(t_i < L_i < t_i + dt, T_i > L_i)$$
$$= g(t_i) S(t_i) \, dt \quad \text{(because $L_i$ and $T_i$ are independent).}$$

Similarly,

$$P(t_i < y_i < t_i + dt, \delta_i = 1) = P(t_i < T_i < t_i + dt, T_i \leq L_i)$$
$$= f(t_i) G(t_i) \, dt \quad \text{(because $L_i$ and $T_i$ are independent).}$$

Because $\delta_i$ is 0 or 1, the previous two expressions can be combined into a single expression,

$$P(t_i < y_i < t_i + dt, \delta_i) = \left[ f(t_i) G(t_i) \right]^{\delta_i} \left[ g(t_i) S(t_i) \right]^{1-\delta_i} \, dt, \quad i = 1, \ldots, n.$$  

Thus the likelihood function of the observations is

$$L = \prod_{i=1}^{n} \left[ f(t_i) G(t_i) \right]^{\delta_i} \left[ g(t_i) S(t_i) \right]^{1-\delta_i}$$
$$= \left( \prod_{i=1}^{n} G(t_i)^{\delta_i} g(t_i)^{1-\delta_i} \right) \left( \prod_{i=1}^{n} f(t_i)^{\delta_i} S(t_i)^{1-\delta_i} \right).$$

Since $G(t_i)$ and $g(t_i)$ do not involve the parameters of interest, $\theta_i$, then the first term can be ignored when obtaining the maximum likelihood solution of the likelihood equation. The likelihood function then reduces to

$$L \propto \prod_{i=1}^{n} f(t_i)^{\delta_i} S(t_i)^{1-\delta_i}.$$

(2.3.1)

where $t_i$ are possible realizations of $Y_i$. Note that the likelihood (2.3.1) involves only the density function, $f$, and the survivor function, $S$, of the lifetimes. The censoring mechanism comes into the likelihood through the values of $\delta_i$. 
CHAPTER 3
THE MIXTURE MODEL

3.1. Introduction and Objectives

In classical discriminant analysis, the categorized data, which is typically called the training sample, is used to construct a rule for assigning future observations from the mixture to one of the underlying categories. There are many situations where discrimination is required between two or more component populations on the basis of some variables or covariates. In medicine, the presence or absence of some symptoms or characteristics are used to categorize patients into several possible diagnoses. For example, blood samples are assayed from two groups of women for detecting potential hemophilia A carriers [Ratnoff and Jones (1977)]. The first group consists of women who do not carry the hemophilia gene, and the second group consists of the obligatory carriers who are women known to be hemophilia A carriers (daughters of hemophiliacs, mothers with more than one hemophilic son, or mothers with one hemophilic son and other hemophilic relatives). The AHF activity and the AHF-like antigen are recorded from the assayed blood to construct a classification procedure for detecting potential carriers. In financial business, variables such as income, age, number of credit cards, credit history and family size may be used to discriminate between people with good and poor credit risks.

In classical discriminant analysis, classification rules are developed based on some measured characteristics or covariates of selected objects known to come from each of the component populations that are examined for differences. The general approach for classification is to construct an optimal linear combination of the covariates and thus transforming a multivariable problem to a univariate one. The assignment of a

-36-
new observation is then based simply on the value of the linear combination for that observation. This particular statistical procedure was first introduced by R.A. Fisher (1936) as a statistical technique in taxonomic problems. Discrimination methods of this kind have been used widely in other areas of statistical analyses since then.

In survival data, however, in addition to the covariates and the exact survival times of the observations being studied, the training sample may consist of uncategorized observations in the form of censored survival times. Classical discriminant analysis cannot take these censored data into account. On the other hand, though finite mixture methods in survival data analysis can include both the censored and non-censored data, traditionally they have not included the covariates. In general finite mixture methods utilize either survival data or covariates alone. By incorporating these covariates into the mixture model, the parameters of the component populations in the training sample can be estimated in terms of the covariates, and thus a classification rule for assigning a new observation with certain covariates into one of the component populations may be obtained. Both the non-censored and censored survival times can be included in the model. So far no work has been pursued in incorporating covariates in the mixture survival data analysis.

The inclusion of the covariates in survival data has been used as a standard analysis to explain the sources of heterogeneity among observations such as in the Kaplan-Meier method or Cox's regression model. Finite mixture distributions have also been used to model heterogeneous survival data. By combining these two approaches, it is hoped that the heterogeneity among observations can be better captured and explained. Furthermore, one can classify an observation into one of the component populations on the basis of the covariates of that particular observation.
3.2. The Mixture Model

The survival distribution of \( M \) observations in the study population is considered to be a mixture of two component distributions with proportions \( \pi_1 \) and \( \pi_2 \), respectively, and where \( \pi_1 \leq \pi_2 \). When each component population has some censored observations, then the total sample will actually consist of 4 component populations: \( M_1 \) observations from group 1 who are not censored (subgroup \( C_1 \)), \( M_2 \) observations from group 2 who are not censored (subgroup \( C_2 \)), \( M_3 \) observations from group 1 who are censored (subgroup \( C_3 \)), and \( M_4 \) observations from group 2 who are censored (subgroup \( C_4 \)). In this situation, the contribution of the non-censored observations to the log-likelihood is in the form of the death density function, and that of the censored observations is in the form of the survival distribution. All four component populations are completely labeled, i.e. have known group memberships at the time of the analysis. Therefore, the log-likelihood of this sample, which is adapted from the log-likelihood of Type 4 by Redner and Walker (1984) can be written as:

\[
LL(\Phi) = \sum_{k=1}^{M_1} \log [p_1 f_1(t_{k1} | \theta_1)] + \sum_{k=1}^{M_2} \log [p_2 f_2(t_{k2} | \theta_2)] + \sum_{k=1}^{M_3} \log [p_3 S_1(t_{k1} | \theta_1)] + \sum_{k=1}^{M_4} \log [p_4 S_2(t_{k2} | \theta_2)] + C.
\] (3.2.1)

where \( M = M_1 + M_2 + M_3 + M_4 \); \( p_i = \frac{M_i}{M} \) and is the proportion of observations in subgroup \( C_i \) (i=1,2,3,4); \( f_1(t_{i1} | \theta_1) \) is the death density function of observations in subgroup \( C_1 \); \( f_2(t_{i2} | \theta_2) \) is the death density function of observations in subgroup \( C_2 \); \( S_1(t_{i1} | \theta_1) \) is the survival distribution of observations in subgroup \( C_3 \); \( S_2(t_{i2} | \theta_2) \) is the survival distribution of observations in subgroup \( C_4 \); \( \Phi = \{ \theta_1, \theta_2 \} \) is the unknown vectors of
parameters in groups 1 and 2; and \( C = \log \left\{ \frac{M_1!}{M_1^!} \frac{M_2!}{M_2^!} \frac{M_3!}{M_3^!} \frac{M_4!}{M_4^!} \right\} \). When the survival times are exponentially distributed, the log-likelihood can then be expressed as:

\[
LL(\Phi) = \sum_{k=1}^{M_1} \log[p_1 \lambda_{k1} \exp \{-\lambda_{k1} t_{k1}\}] + \sum_{k=1}^{M_2} \log[p_2 \lambda_{k2} \exp \{-\lambda_{k2} t_{k2}\}] + \sum_{k=1}^{M_3} \log[p_3 \exp \{-\lambda_{k1} t_{k1}\}] + \sum_{k=1}^{M_4} \log[p_4 \exp \{-\lambda_{k2} t_{k2}\}],
\]

(3.2.2)

where \( \lambda_{k1} \) and \( \lambda_{k2} \) are the hazard functions for the \( k \)th observation in group 1 or 2, respectively. The constant \( C \) has been dropped from the above log-likelihood function since its value does not affect the estimation procedure. The exponential distribution is used here because it provides an excellent fit to a variety of clinical survival data and it is the most easily used as well.

The covariates are incorporated to the mixture model through the hazard function of each component. If the log-linear exponential hazard model is considered and there are \( u \) covariates measured from each observation, then the \( k \)th observation in groups 1 and 2 have the following hazard functions, respectively:

\[
\lambda_{k1} = \exp(\sum_{j=0}^{u} b_{j1} z_{kj})
\]

and

\[
\lambda_{k2} = \exp(\sum_{j=0}^{u} b_{j2} z_{kj}),
\]

where \( z_{k0} \) is set to one for all patients. The term \( b_0 \) represents the underlying hazard rate or the intercept. Substituting the above hazard function in the log-likelihood function yields:
\[ \text{LL}(\Phi) = \sum_{k=1}^{M_1} \log \left\{ p_1 \left[ \exp \left( \sum_{j=0}^{u} b_{1j} z_{kj} \right) \right] \exp \left\{ -\left[ \exp \left( \sum_{j=0}^{u} b_{1j} z_{kj} \right) \right] t_{k1} \right\} \right\} + \]

\[ \sum_{k=1}^{M_2} \log \left\{ p_2 \left[ \exp \left( \sum_{j=0}^{u} b_{2j} z_{kj} \right) \right] \exp \left\{ -\left[ \exp \left( \sum_{j=0}^{u} b_{2j} z_{kj} \right) \right] t_{k2} \right\} \right\} + \]

\[ \sum_{k=1}^{M_3} \log \left\{ p_3 \exp \left\{ -\left[ \exp \left( \sum_{j=0}^{u} b_{1j} z_{kj} \right) \right] t_{k1} \right\} \right\} + \]

\[ \sum_{k=1}^{M_4} \log \left\{ p_4 \exp \left\{ -\left[ \exp \left( \sum_{j=0}^{u} b_{2j} z_{kj} \right) \right] t_{k2} \right\} \right\}. \tag{3.2.3} \]

The unknown mixture parameters to solve for are written in terms of the covariate coefficients denoted by \( \Phi \). The maximum likelihood estimate of \( \Phi \) is the solution to the following system of equations:

\[ \frac{\partial \text{LL}(\Phi)}{\partial \Phi} = 0. \tag{3.2.4} \]

The Newton-Raphson iterative method is used to obtain the approximate solutions to equation (3.2.4). In general, the recursive step of the algorithm is defined by:

\[ \Phi^{(s+1)} = \Phi^{(s)} - \left[ D^2 \text{LL}(\Phi^{(s)}) \right]^{-1} \text{DLL}(\Phi^{(s)}), \quad s = 0, 1, \ldots, \tag{3.2.5} \]

where \( \text{DLL}(\Phi) = \frac{\partial \text{LL}(\Phi)}{\partial \Phi} \) and \( D^2 \text{LL}(\Phi) = \frac{\partial^2 \text{LL}(\Phi)}{\partial \Phi^2} \). For this particular case, \( \Phi = \{b_{01}, b_{11}, b_{21}, \ldots, b_{u1}, b_{02}, b_{12}, b_{22}, \ldots, b_{u2}\} \), since the proportions of all component populations \( p_i \)'s are known.

The first derivatives of equation (3.2.3) with respect to the unknown parameters in groups 1 and 2 are, respectively:

\[ \frac{\partial \text{LL}(\Phi)}{\partial b_{j1}} = \sum_{k=1}^{M_1} \left[ z_{kj} - z_{kj} t_{k1} \lambda_{k1} \right] - \sum_{k=1}^{M_2} \left[ z_{kj} t_{k1} \lambda_{k1} \right]. \quad j = 0, 1, 2, \ldots, u \]
and
\[
\frac{\partial L^2(\Phi)}{\partial b_{j2}} = \sum_{k=1}^{M_2} \left[ z_{kj} - z_{tk2} \right] \lambda_{k1} - \sum_{k=1}^{M_4} \left[ z_{kj} t_{k2} \lambda_{k2} \right], \quad j = 0, 1, \ldots, u,
\]

where \( \lambda_{k1} \) and \( \lambda_{k2} \) are as defined previously. The second derivatives of equation (3.2.3) with respect to the unknown parameters in groups 1 and 2 are, respectively:

\[
\frac{\partial L^2(\Phi)}{\partial b_{j1} \partial b_{q1}} = -\sum_{k=1}^{M_1} \left[ z_{kj} z_{kq} t_{k1} \lambda_{k1} \right] - \sum_{k=1}^{M_3} \left[ z_{kj} z_{kq} t_{k1} \lambda_{k2} \right], \quad j, q = 0, 1, \ldots, u,
\]

and

\[
\frac{\partial L^2(\Phi)}{\partial b_{j2} \partial b_{q2}} = -\sum_{k=1}^{M_2} \left[ z_{kj} z_{kq} t_{k2} \lambda_{k1} \right] - \sum_{k=1}^{M_4} \left[ z_{kj} z_{kq} t_{k2} \lambda_{k2} \right], \quad j, q = 0, 1, \ldots, u.
\]

In general, when there are \( u \) covariates, the equation has to be solved for \( 2(u+1) \) unknown parameters.

The initial starting vector \( \Phi^{(0)} \) is obtained through the relationship between the survival distribution and the covariates as follows:

\[
S_i(t) = \exp(-\lambda_i t)
\]

\[
= \exp\{-\exp(b_{0i} + b_{1i}z_{1i} + b_{2i}z_{2i} + \ldots + b_{ui}z_{ui}) t\}, \quad i = 1, 2.
\]

Taking the natural logarithm twice of both sides of the above equation yields:

\[
\log\left( \frac{-\log[S_i(t)]}{t} \right) = b_{0i} + b_{1i}z_{1i} + b_{2i}z_{2i} + \ldots + b_{ui}z_{ui}
\]

for \( i = 1 \) and 2. By letting \( \log\left( \frac{-\log[S_i(t)]}{t} \right) = y_i \), the equation becomes a simple multiple linear regression of \( y_i \) (\( i = 1, 2 \)) on the covariates:

\[
y_i = b_{0i} + b_{1i}z_{1i} + b_{2i}z_{2i} + \ldots + b_{ui}z_{ui}.
\]
The least squares estimates of the regression coefficients are obtained separately for groups 1 and 2. The empirical survival distribution \( \hat{S}_i(t) \) can be estimated from the data through the procedures described in Gross and Clark (1975):

1. in each group, rank all the survival times from the smallest to the largest, regardless of whether they are censored or not as: \( t_{(1)} \leq t_{(2)} \leq \cdots \leq t_{(n)} \),

and

2. estimate the empirical survival distribution for the non-censored times \( t_{(k)} \) only as \( \hat{S}_i(t_{(k)}) = \frac{n_i - r_{ki} + 1}{n_i + 1} \) \( (i=1,2) \), where \( n_i \) is the total sample in group \( i \) and \( r_{ki} \) is the rank of \( t_{(k)} \) in group \( i \).

As a simple example, suppose the recorded survival times are: \( 2, 5, 1, 3^+, 1.5, 4^+, 6, 5.5 \), with a plus sign indicating the censored data. After these survival times are ordered and ranked, the following table can be obtained.

<table>
<thead>
<tr>
<th>( t_k )</th>
<th>Rank ( (r) )</th>
<th>( \hat{S}(t_k) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0.89=8/9</td>
</tr>
<tr>
<td>1.5+</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>0.67=6/9</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>0.56=5/9</td>
</tr>
<tr>
<td>4+</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>0.33=3/9</td>
</tr>
<tr>
<td>5.5</td>
<td>7</td>
<td>0.22=2/9</td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>0.11=1/9</td>
</tr>
</tbody>
</table>

These procedures to obtain the empirical survival rate should be performed separately in groups 1 and 2.
3.3. Comparisons of Models

In this section, the proposed mixture model with the covariates is compared to the one without the covariates and to the survival distribution obtained from the life table method through graphical representations. Based on equation (3.2.2), the log-likelihood of the mixture model without the covariates is simply:

\[
LL(\Phi) = \sum_{k=1}^{M_1} [\log p_1 + \log \lambda_1 - \lambda_1 t_{k1}] + \sum_{k=1}^{M_2} [\log p_2 + \log \lambda_2 - \lambda_2 t_{k2}] - \\
\sum_{k=1}^{M_3} [\log p_3 + \lambda_1 t_{k1}] - \sum_{k=1}^{M_4} [\log p_4 + \lambda_2 t_{k2}].
\] (3.2.6)

The above log-likelihood function has to be solved for \( \lambda_1 \) and \( \lambda_2 \), which are the hazard rates for groups 1 and 2, respectively. The Newton-Raphson iterative procedure (3.2.5) is used to obtain the estimate of the unknown hazard rates. The first derivatives of the log-likelihood (3.2.6) with respect to the unknown parameters in groups 1 and 2 are, respectively:

\[
\frac{\partial LL(\Phi)}{\partial \lambda_1} = \sum_{k=1}^{M_1} \left\{ \frac{1}{\lambda_1} - t_{k1} \right\} - \sum_{k=1}^{M_2} t_{k1}
\]

and

\[
\frac{\partial LL(\Phi)}{\partial \lambda_2} = \sum_{k=1}^{M_2} \left\{ \frac{1}{\lambda_2} - t_{k2} \right\} - \sum_{k=1}^{M_4} t_{k2}.
\]

The second derivatives of the log-likelihood in (3.2.6) with respect to the unknown parameters in groups 1 and 2 are, respectively:

\[
\frac{\partial LL^2(\Phi)}{\partial \lambda_1^2} = - \sum_{k=1}^{M_1} \left\{ \frac{1}{\lambda_1^2} \right\}
\]
\[
\frac{\partial \text{LL}^2(\Phi)}{\partial \lambda_2^2} = - \sum_{k=1}^{M_2} \left\{ \frac{1}{\lambda_2^2} \right\}.
\]

The initial starting vector \([\lambda_1^{(0)}, \lambda_2^{(0)}]\) is obtained through the relationship between the survival distribution and the hazard rate, \(S_i(t) = \exp(-\lambda_i t)\), for \(i = 1, 2\). Taking the natural logarithm of both sides of this equation yields: \(-\log[S_i(t)] = \lambda_i t\), and by letting \(-\log[S_i(t)] = y_i\), the equation becomes a simple linear regression of \(y_i\) on the survival times \(y_i = \lambda_i t\) \((i = 1, 2)\). The least squares estimates of the regression coefficients which estimate the hazard rates are obtained separately for groups 1 and 2. The empirical survival distribution \(\hat{S}_i(t)\) can be estimated from the data through the procedures described in section 3.2.

After obtaining the hazard rates, the component and mixture survival distributions based on these estimated hazard rates are calculated. In general, graphical representations of the two component and the mixture survival distributions can be obtained from \(\hat{S}_i(t; \lambda_i) = \frac{1}{n_1} \sum_{k=1}^{n_1} \hat{S}_i(t_k; \lambda_i)\) for \(i = 1\) and 2, and \(\hat{S}(t; \lambda) = \pi_1 \hat{S}_1(t; \lambda_1) + \pi_2 \hat{S}_2(t; \lambda_2)\), where \(n_1\) and \(n_2\) are the number of observations in groups 1 and 2, respectively. The estimate of the survival distributions of the training sample can be calculated using the life table method. The survival distributions obtained from the mixture models with and without the covariates are graphed and their fits to the life table survival can be compared visually.

3.4. Classification Procedures

After the coefficients of the covariates \(\Phi\) are obtained, a probability rule based on these estimated parameters must be constructed to classify a new observation
with certain covariates into one of the component populations. Five different classification methods are proposed and compared in this section. Furthermore, the performances of the proposed classification procedures are compared to the neighborhood classification procedure, which is a non-parametric discrimination procedure. The neighborhood procedure is chosen instead of the Fisher's discriminant procedure because unlike the latter, the neighborhood method does not require that the component populations have a multivariate normal distribution. The non-parametric form assumption of the distribution is important here because the proposed mixture model is constructed based on exponential distributions, and not on normal distributions.

**Classification Procedure 1.**

This classification procedure is based on two computed hazard values for the observation that has to be classified. When the new observation has the covariates \(\{z_1^*, z_2^*, \ldots, z_u^*\}\), the two possible hazard rates are computed as \(\lambda_1^* = \exp\left(\sum_{j=0}^{n} \hat{b}_{j1} z_j^*\right)\), or \(\lambda_2^* = \exp\left(\sum_{j=0}^{n} \hat{b}_{j2} z_j^*\right)\), assuming that such an observation comes from group 1 or 2, respectively. Depending on the values of \(\lambda_1^*\) and \(\lambda_2^*\), this observation must be assigned to either group 1 or 2 based on the higher probability that its true hazard rate is \(\lambda_1^*\) or \(\lambda_2^*\), respectively. Let \(\Lambda_1\) and \(\Lambda_2\) be the random variables describing the hazard rates of groups 1 and 2, respectively. If the new observation is more likely to belong to group 1, then the right tail probability under group 1, \(P(\Lambda_1 \geq \lambda_1^*)\), should be larger than the left tail probability under group 2, \(P(\Lambda_2 \leq \lambda_2^*)\). If the new observation is more likely to belong to group 2, then \(P(\Lambda_1 \geq \lambda_1^*)\) should be smaller than \(P(\Lambda_2 \leq \lambda_2^*)\). These probabilities are estimated from the empirical hazard distributions of the training sample. Figure 3.1 illustrates this classification procedure.
The procedures to obtain this probability from the training sample in group 1 can be described as follows: (1) order the hazard rates in group 1 from the smallest to the largest to obtain $\lambda_{(1)} \leq \lambda_{(2)} \leq \ldots \leq \lambda_{(n_1)}$, where the first subscript indicates the rank, the second subscript indicates group 1, and $n_1$ is the number of observations in group 1; and (2) for a new $\lambda_1^{*}$ in group 1, find the hazard rate $\lambda_{(k)}$ in the training sample such that $\lambda_{(k)} \geq \lambda_1^{*}$, and then compute the right tail probability as $P(\Lambda_1 \geq \lambda_1^{*}) = 1 - P(\Lambda_1 < \lambda_{(k)}) = 1 - \frac{k-1}{n_1} = \frac{n_1 - k + 1}{n_1}$. 

Figure 3.1. Distribution of individuals hazard rates for groups 1 and 2
For example, in the following table, the hazard rates and their ranks from the training sample in group 1 are obtained.

<table>
<thead>
<tr>
<th>Rank (k)</th>
<th>$\lambda_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.04</td>
</tr>
<tr>
<td>2</td>
<td>0.04</td>
</tr>
<tr>
<td>3</td>
<td>0.05</td>
</tr>
<tr>
<td>4</td>
<td>0.055</td>
</tr>
<tr>
<td>5</td>
<td>0.06</td>
</tr>
</tbody>
</table>

When $\lambda_1^* = 0.045$, the value of the hazard rate from the training sample that is greater than or equal to 0.045 is 0.05 with $k=3$, therefore $P(A_1 \geq 0.045) = \frac{5-3+1}{5} = 0.60$.

When $\lambda_1^* = 0.04$, the value of the hazard rate from the training sample that is greater than or equal to 0.04 is 0.04 with $k=1$, therefore $P(A_1 \geq 0.04) = \frac{5-1+1}{5} = 1$.

Similarly, the procedures to obtain this probability from the training sample in group 2 can be described as follows: (1) order the hazard rates in group 2 from the smallest to the largest to obtain $\lambda_{(1)}^2 \leq \lambda_{(2)}^2 \leq \ldots \leq \lambda_{(n_2)}^2$, where the first subscript indicates the rank, the second subscript indicates group 2, and $n_2$ is the number of observations in group 2; and (2) for a new $\lambda_2^*$ in group 2, find the hazard $\lambda_{(k)}^2$ such that $\lambda_{(k)}^2 > \lambda_2^*$, and then compute the left tail probability as $P(A_2 \leq \lambda_2^*) = P(A_2 < \lambda_{(k)}^2) = \frac{k-1}{n_2}$. For example, in the following table, the hazard rates and their ranks in group 2 are obtained.
Table 1.3. Hazard rates for group 2.

<table>
<thead>
<tr>
<th>Rank (k)</th>
<th>$\lambda_2^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.045</td>
</tr>
<tr>
<td>2</td>
<td>0.06</td>
</tr>
<tr>
<td>3</td>
<td>0.065</td>
</tr>
<tr>
<td>4</td>
<td>0.07</td>
</tr>
<tr>
<td>5</td>
<td>0.08</td>
</tr>
<tr>
<td>6</td>
<td>0.089</td>
</tr>
<tr>
<td>7</td>
<td>0.09</td>
</tr>
</tbody>
</table>

When $\lambda_2^* = 0.065$, the value of the hazard rate from the training sample that is strictly greater than 0.065 is 0.07 with $k=4$, therefore $P(\Lambda_2 \leq 0.065) = \frac{4-1}{7} = 0.43$. When $\lambda_2^* = 0.075$, the value of the hazard rate from the training sample that is strictly greater than 0.075 is 0.08 with $k=5$, therefore $P(\Lambda_2 \leq 0.075) = \frac{5-1}{7} = 0.57$.

A probability rule for classifying the new observation into one of the groups can then be constructed as:

if $P(\Lambda_1^* \geq \lambda_1^*) > P(\Lambda_2^* \leq \lambda_2^*)$, then classify in group 1 or

if $P(\Lambda_1^* \geq \lambda_1^*) \leq P(\Lambda_2^* \leq \lambda_2^*)$, then classify in group 2.

Classification Procedure 2.

This classification procedure combines the first classification procedure with the Bayes' rule. The problem here is estimating the probability of being in group 1 given $\lambda_1^*$, i.e. $\Pr(\text{group 1} \mid \lambda_1^*)$, and the probability of being in group 2 given $\lambda_2^*$, i.e. $\Pr(\text{group 2} \mid \lambda_2^*)$. After $\lambda_1^*$ and $\lambda_2^*$ are computed, the following probabilities are obtained from the training sample through the same procedures described in
classification procedure 1: \( P(\Lambda_1 \geq \lambda_1^* \mid \lambda_1^* \in \text{group 1}) \), \( P(\Lambda_2 \leq \lambda_1^* \mid \lambda_1^* \in \text{group 2}) \),
\( P(\Lambda_1 \geq \lambda_2^* \mid \lambda_2^* \in \text{group 1}) \), and \( P(\Lambda_2 \leq \lambda_2^* \mid \lambda_2^* \in \text{group 2}) \). The posterior probabilities of the group memberships can then be calculated from:

\[
Pr(\text{group 1} \mid \lambda_1^*) = \frac{P(\Lambda_1 \geq \lambda_1^* \mid \lambda_1^* \in \text{group 1}) \pi_1}{P(\Lambda_1 \geq \lambda_1^* \mid \lambda_1^* \in \text{group 1}) \pi_1 + P(\Lambda_2 \leq \lambda_1^* \mid \lambda_1^* \in \text{group 2}) \pi_2}
\]

and

\[
Pr(\text{group 2} \mid \lambda_2^*) = \frac{P(\Lambda_2 \leq \lambda_2^* \mid \lambda_2^* \in \text{group 2}) \pi_2}{P(\Lambda_1 \geq \lambda_2^* \mid \lambda_2^* \in \text{group 1}) \pi_1 + P(\Lambda_2 \leq \lambda_2^* \mid \lambda_2^* \in \text{group 2}) \pi_2}.
\]

The classification rule is constructed as:

- if \( Pr(\text{group 1} \mid \lambda_1^*) > Pr(\text{group 2} \mid \lambda_2^*) \), then classify in group 1 or
- if \( Pr(\text{group 1} \mid \lambda_1^*) \leq Pr(\text{group 2} \mid \lambda_2^*) \), then classify in group 2.

**Classification Procedure 3.**

This classification procedure takes into account the survival time and censorship status of the new observation and the fact that this observation will have a mixture hazard rate. If the observation is not censored, then the observation must be from either subgroup \( C_1 \) or \( C_2 \), otherwise the observation must be from either subgroup \( C_3 \) or \( C_4 \). The death density function or the survival rate of a new observation with survival time \( t^* \) and covariates \( \{z_1^*, z_2^*, ..., z_u^*\} \) in each component population can be calculated from the estimated covariate coefficients. If the observation is not censored then the mixture death density function based on the prior probabilities of group memberships is

\[
f(t^*) = p_1 f_1(t^*) + p_2 f_2(t^*),
\]

where \( f_1(t^*) = \lambda_1^* \exp(-\lambda_1^* t^*) \)

and \( f_2(t^*) = \lambda_2^* \exp(-\lambda_2^* t^*) \). If the observation is censored, then the mixture survival rate based on the prior probabilities of group memberships is
\[ S(t^*) = p_3 S_1(t^*) + p_4 S_2(t^*), \]
where \( S_1(t^*) = \exp(-\lambda_1^* t^*) \) and \( S_2(t^*) = \exp(-\lambda_2^* t^*). \) When the new observation is not censored, the posterior probability that this observation belongs to group 1 or 2 is [McLachlan and Basford (1988)]:

\[
p_1(t^*) = \frac{p_1 f_1(t^*)}{f(t^*)} \quad \text{or} \quad p_2(t^*) = \frac{p_2 f_2(t^*)}{f(t^*)}.
\]

Similarly, when the new observation is censored the posterior probability that this observation belongs to group 1 or 2 is:

\[
p_1(t^*) = \frac{p_3 S_1(t^*)}{S(t^*)} \quad \text{or} \quad p_2(t^*) = \frac{p_4 S_2(t^*)}{S(t^*)}.
\]

Since the new observation comes from the mixture population at time \( t^* \), the new updated mixture death density and the survival rates at time \( t^* \), based on the posterior probabilities of group memberships are, respectively:

\[
f_{\text{mix}}(t^*) = p_1(t^*) f_1(t^*) + p_2(t^*) f_2(t^*),
\]

and

\[
S_{\text{mix}}(t^*) = p_1(t^*) S_1(t^*) + p_2(t^*) S_2(t^*).
\]

The mixture hazard rate of this new observation is then:

\[
\lambda(t^*) = \frac{f_{\text{mix}}(t^*)}{S_{\text{mix}}(t^*)}.
\]

For a non-censored observation, \( p_1(t^*) \) and \( p_2(t^*) \) from (3.3.1) should be used. Otherwise \( p_1(t^*) \) and \( p_2(t^*) \) from (3.3.2) should be used. The right and left tail probabilities of \( \lambda(t^*) \) are then estimated from the empirical hazard distributions as described in classification procedure 1. The classification rule is:

- If \( \{\Lambda_1 \geq \lambda(t^*)\} > \{\Lambda_2 \leq \lambda(t^*)\} \), then classify in group 1 or
- If \( \{\Lambda_1 \geq \lambda(t^*)\} \leq \{\Lambda_2 \leq \lambda(t^*)\} \), then classify in group 2.
Classification Procedure 4.

This classification procedure compares the appropriate posterior probabilities of group memberships at time $t^*$, i.e. $p_1(t^*)$ and $p_2(t^*)$ that are computed from equations (3.3.1) and (3.3.2) in classification procedure 3. The classification rule is:

if $p_1(t^*) > p_2(t^*)$, then classify in group 1 or

if $p_1(t^*) \leq p_2(t^*)$, then classify in group 2.

Classification Procedure 5.

This classification procedure adapts the non-parametric neighborhood method, constructed from the Mahalanobis distance which is based on the sample variance. The mixture hazard rate $\lambda_m^*$ for an observation with covariates $(z_1^*, \ldots, z_u^*)$ is obtained as in classification procedure 3. This procedure uses the assumption that the component populations have the same variances, therefore the overall variance of the training sample hazard is computed by pooling the variances of the hazards in the two components as:

$$V_p = \frac{(n_1 - 1) V_1 + (n_2 - 1) V_2}{n_1 + n_2 - 2},$$

where $V_i$ is the variance of the hazard values in group $i$ ($i = 1, 2$) and $n_i$ is the number of observations in group $i$ ($i = 1, 2$). The idea here is to find the "minimum distance" between the new mixture hazard value and all the training sample hazard values in groups 1 and 2. These distances are, respectively:

$$d_{k1} = (\lambda_{k1} - \lambda_m^*) (V_p)^{-1} (\lambda_{k1} - \lambda_m^*'),$$

for $k=1, 2, \ldots, n_1$, and

$$d_{k2} = (\lambda_{k2} - \lambda^*) (V_p)^{-1} (\lambda_{k2} - \lambda^*'),$$

for $k=1, 2, \ldots, n_2$. Obtain the minimum of the distances in groups 1 and 2 as
\[ d_i = \min\{d_{ki}, k=1,2, \ldots, n_i\} \text{ for } i = 1, 2. \] The classification rule is based on the minimum of \( d_1 \) and \( d_2 \):

if \( d_1 < d_2 \), then classify in group 1, or

if \( d_1 \geq d_2 \), then classify in group 2.

**The Neighborhood Classification Procedure.**

The non-parametric neighborhood method computes the distance between the covariates of the new observation and the covariates in the training sample. The training sample consists of \( n_1 \) observations of the covariates \( Z'_1 = [z_{11}, z_{21}, \ldots, z_{n_11}] \) in group 1, and \( n_2 \) observations of the covariates \( Z'_2 = [z_{12}, z_{22}, \ldots, z_{n_22}] \) in group 2. From these data, the sample mean vectors and covariance matrices are determined by

\[
\bar{\mathbf{z}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} z_{ji} \quad \text{and} \quad \mathbf{V}_i = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (z_{ji} - \bar{\mathbf{z}}_i)(z_{ji} - \bar{\mathbf{z}}_i)', \quad \text{for } i=1,2.
\]

The pooled variance-covariance matrix can be obtained as:

\[
\mathbf{V}_p = \frac{(n_1 - 1) \mathbf{V}_1 + (n_2 - 1) \mathbf{V}_2}{n_1 + n_2 - 2}.
\]

The distances of the new set of covariates \( \mathbf{z}^* = [z_1^*, z_2^*, \ldots, z_u^*]' \) from all pairs of covariates in groups 1 and 2 can be computed from, respectively:

\[
d_{k1} = (\bar{\mathbf{z}}_1 - \mathbf{z}^*)'(\mathbf{V}_p)^{-1}(\bar{\mathbf{z}}_1 - \mathbf{z}^*),
\]

for \( k = 1,2, \ldots, n_1 \), and

\[
d_{k2} = (\bar{\mathbf{z}}_2 - \mathbf{z}^*)'(\mathbf{V}_p)^{-1}(\bar{\mathbf{z}}_2 - \mathbf{z}^*),
\]

for \( k = 1,2, \ldots, n_2 \). Obtain the minimum of the distances in groups 1 and 2 as

\[ d_i = \min\{d_{ki}, k=1,2, \ldots, n_i\} \text{ for } i = 1, 2. \] The classification rule is based on the
minimum of \( d_1 \) and \( d_2 \):

\[
\begin{align*}
\text{if } d_1 < d_2, \text{ then classify in group 1, or} \\
\text{if } d_1 \geq d_2, \text{ then classify in group 2.}
\end{align*}
\]

In all of the classification procedures described in this section, in case of a tie in the classification rules, the decision is always to classify the new observation in group 2 instead of group 1, because the proportion of observations in group 2, \( \pi_1 \), is greater than or equal to the proportion of observations in group 2, \( \pi_2 \).

Even though the proposed model does not have any restrictions on the number of covariates that can be included in the model, it is recommended that a screening of the covariates be performed first. More covariates will result in more parameters to estimate, and thus more equations to solve, making the numerical computations more complex. Covariate screening may be conducted univariately or multivariately using the classical survival analysis method such as the Kaplan-Meier or Cox’s regression model. Previous experience or well-known facts about the covariates and the data at hand may also be useful in deciding which covariates to use.

To illustrate the methodology, data from a prospective breast cancer clinical trial based at Case Western Reserve University will be used in chapter 4. The classification procedures will also be studied under different situations by applying the methodology to simulated clinical trial data in chapter 5.
CHAPTER 4
APPLICATION

4.1. The Data

In this chapter, the models are illustrated using data from a prospective adjuvant clinical trial in breast cancer based at Case Western Reserve University, Cleveland, Ohio. From September 1974 to June 1979, 311 stage II breast cancer patients were entered into the trial. Inclusion criteria were no previous history of breast cancer or other primary tumor and age less than 76 years. All patients received a modified radical mastectomy with complete axillary lymph node dissection and pathologic staging. Patients were stratified by 1) the number of positive axillary lymph nodes: less than four and four or more, and 2) the estrogen receptor (ER) contents of the primary tumor: negative if ER was less than 3 femtomoles/mg of cytosol protein, and positive otherwise. They were then randomized to one of three treatments: 1) chemotherapy consisting of Cytoxan, Methotrexate, 5-Fluorouracil (CMF) for 1 year; 2) chemotherapy with Tamoxifen (CMFT) for 1 year; and 3) CMFT followed by a second year of BCG (bacillus Calmette-Guerin) immunotherapy (CMFTBCG). Number of patients in the CMF, CMFT and CMFTBCG were 99, 102 and 110, respectively. As part of the study, other patient characteristics or covariates were also collected at the time of diagnosis. These covariates were age at time of mastectomy, menopausal status, tumor diameter, family history of breast cancer, weight and race.

A regular follow-up on these patients has been maintained by the Breast Cancer Research Project Office at Case Western Reserve University. In this follow-up, recurrence and the survival status of the patients are updated. The data base are kept in the University Hospitals of Cleveland IBM computer mainframe. Hubay et al. (1985)
and Crowe, et al. (1990) have published the results of this study. The follow-up data that will be used in this application was updated in October 1991.

To illustrate the model, these patients are considered as a mixture of those with axillary positive nodes (PN) between 1 and 3 (group 1) and those with PN $\geq 4$ (group 2). This situation may not be very practical, because most of the time the number of positive lymph nodes are known, however it will serve as a good example since the distributions of the two groups are well separated. In this data, about 47% of the total sample comes from group 1, i.e. $\pi_1=0.47$ and 53% from group 2, i.e. $\pi_2=0.53$.

In this example, analyses using recurrence and death from breast cancer as the end points will be considered. The recurrence-free interval is measured as the interval in years between the mastectomy date and a documented date of a recurrence as of the last follow-up. Patients who have not recurred are considered censored at their last date of follow-up. At the time of the analysis, group 1 consists of $M_1=64$ patients who have recurred, and $M_3=82$ patients who have not recurred. Group 2 consists of $M_2=124$ patients who have recurred, and $M_4=41$ patients who have not recurred.

Breast cancer survival is measured as the interval in years between the mastectomy date and time of death from breast cancer. Patients who have not died from breast cancer are considered censored at their last date of follow-up. At the time of the analysis, group 1 consists of $M_1=67$ patients who died from breast cancer, and $M_3=79$ patients who did not die from breast cancer. Group 2 consists of $M_2=119$ patients who died from breast cancer, and $M_4=46$ patients who did not die from breast cancer.

To see whether the covariates, the number of recurrences and breast cancer deaths are statistically different in groups 1 and 2, Fisher's exact test is conducted when the variables are of categorical type, and the two-sample $t$-test is performed when the variables are of continuous type. The results are presented in the following table.
Table 4.1. Distribution of the covariates, recurrences and breast cancer deaths in groups 1 and 2.

<table>
<thead>
<tr>
<th>COVARIATE</th>
<th>GROUP 1a</th>
<th>GROUP 2b</th>
<th>P-valuec</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estrogen Receptor (ER)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>–</td>
<td>33 (23%)</td>
<td>44 (27%)</td>
<td>0.43</td>
</tr>
<tr>
<td>+</td>
<td>113 (77%)</td>
<td>121 (73%)</td>
<td></td>
</tr>
<tr>
<td>Tumor Diameter (cm.)</td>
<td>3.2 ± 1.0</td>
<td>3.9 ± 2.5</td>
<td>0.01</td>
</tr>
<tr>
<td>(Mean ± Std. Deviation)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td>54 ± 10</td>
<td>53 ± 10</td>
<td>0.52</td>
</tr>
<tr>
<td>(Mean ± Std. Deviation)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race - White</td>
<td>116 (79%)</td>
<td>140 (85%)</td>
<td>0.24</td>
</tr>
<tr>
<td>- Black</td>
<td>30 (21%)</td>
<td>25 (15%)</td>
<td></td>
</tr>
<tr>
<td>Menopausal Status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Premenopausal</td>
<td>46 (32%)</td>
<td>48 (29%)</td>
<td>0.71</td>
</tr>
<tr>
<td>- Postmenopausal</td>
<td>100 (68%)</td>
<td>117 (71%)</td>
<td></td>
</tr>
<tr>
<td>Recurrence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Recurrences</td>
<td>64 (44%)</td>
<td>124 (75%)</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>- Censored</td>
<td>82 (55%)</td>
<td>41 (25%)</td>
<td></td>
</tr>
<tr>
<td>Breast Cancer Survival</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Deaths</td>
<td>54 (37%)</td>
<td>113 (68%)</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>- Censored</td>
<td>92 (63%)</td>
<td>52 (32%)</td>
<td></td>
</tr>
</tbody>
</table>

a PN ≤ 3
b PN ≥ 4
c P-value is obtained from the Fisher’s exact test for categorical variable, or the two-sample t-test for continuous variable.
From Table 4.1 it can be seen that on the average, tumor diameter in group 1 is significantly greater than tumor diameter in group 2 (p=0.01). The distributions of the other covariates in group 1 are not significantly different from those in group 2. There are significantly fewer recurrences (p<0.01) and deaths (p<0.01) in group 1 than in group 2.

To examine how the covariates affect the survival in groups 1 and 2, Kaplan-Meier survival analyses with the grouping (PN) variable as the strata, and ER, tumor diameter, age, race, and menopausal status as the covariates are performed using procedure LIFETEST in SAS [Delong (1985)]. The log-rank tests for comparing the recurrence-free and breast cancer survivals in groups 1 and 2 are both significant (p < 0.01 for both cases). Tables 4.2 and 4.3 present the univariate effect of each of the covariate on recurrence-free survival and breast cancer survival after adjusting for PN variable, obtained from procedure LIFETEST in SAS.

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Test Statistic</th>
<th>Chi-square Statistic</th>
<th>P-value(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ER</td>
<td>10.87</td>
<td>3.83</td>
<td>0.05</td>
</tr>
<tr>
<td>Tumor Diameter</td>
<td>−66.82</td>
<td>5.08</td>
<td>0.02</td>
</tr>
<tr>
<td>Age</td>
<td>169.50</td>
<td>1.57</td>
<td>0.21</td>
</tr>
<tr>
<td>Race</td>
<td>−5.69</td>
<td>1.28</td>
<td>0.26</td>
</tr>
<tr>
<td>Menopausal Status</td>
<td>−15.28</td>
<td>5.69</td>
<td>0.02</td>
</tr>
</tbody>
</table>

\(^a\) P-value is obtained from the log-rank test
Table 4.3. Univariate effect of the covariates on breast cancer survival after controlling for PN variable.

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Test Statistic</th>
<th>Chi-square Statistic</th>
<th>P-value$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ER</td>
<td>12.97</td>
<td>6.36</td>
<td>0.01</td>
</tr>
<tr>
<td>Tumor Diameter</td>
<td>-61.97</td>
<td>4.54</td>
<td>0.03</td>
</tr>
<tr>
<td>Age</td>
<td>136.00</td>
<td>1.13</td>
<td>0.29</td>
</tr>
<tr>
<td>Race</td>
<td>-9.09</td>
<td>3.79</td>
<td>0.05</td>
</tr>
<tr>
<td>Menopausal Status</td>
<td>-10.63</td>
<td>3.09</td>
<td>0.08</td>
</tr>
</tbody>
</table>

$^a$ P-value is obtained from the log-rank test

Tables 4.2 shows that after controlling for PN variable, ER positive is significantly related to longer recurrence-free survival, while larger tumor diameter and postmenopausal status are both significantly related to shorter recurrence-free survival. Table 4.3 shows that ER positive is also related to longer breast cancer survival, while larger tumor diameter and blacks are significantly related to shorter breast cancer survival. For simplicity, only two covariates are used in the mixture model, which are ER and tumor diameter. This choice also seems reasonable considering that it is well known from the breast cancer literature that ER status and tumor diameter play an important role in recurrence and survival from breast cancer.

4.2. The Model

Estrogen receptor status is coded 0 when it is negative and 1 when it is positive. Tumor diameter is used in the model as a continuous variable. If the ER status is denoted by $Z_1$ and tumor diameter by $Z_2$, then from (3.2.2) the log-likelihood
of this sample can be written as:

\[
LL(\Phi) = \sum_{k=1}^{M_1} \log \left[ p_1 \lambda_{k1} \exp \{-\lambda_{k1} t\} \right] + \sum_{k=1}^{M_2} \log \left[ p_2 \lambda_{k2} \exp \{-\lambda_{k2} t\} \right] + \\
\sum_{k=1}^{M_3} \log \left[ p_3 \exp \{-\lambda_{k1} t\} \right] + \sum_{k=1}^{M_4} \log \left[ p_4 \exp \{-\lambda_{k2} t\} \right],
\]

(4.2.1)

where

\[
\lambda_{k1} = \exp(b_{01} + b_{11} z_{k1} + b_{21} z_{k2})
\]

(4.2.2)

and

\[
\lambda_{k2} = \exp(b_{02} + b_{12} z_{k1} + b_{22} z_{k2}).
\]

(4.2.3)

The maximum likelihood estimates \( \hat{\Phi} \) are the solution of the following system of equations:

\[
\frac{\partial LL(\Phi)}{\partial b_{ji}} = 0, \ j=0,1,2; \ i=1,2.
\]

The Newton-Raphson iterative method is used to obtain the approximate solutions to the log-likelihood function as:

\[
\Phi(s+1) = \Phi(s) - \left[ D^2 LL(\Phi(s)) \right]^{-1} DLL(\Phi(s)), \ s = 0, 1, \ldots
\]

where \( D^2 LL(\Phi) = \frac{\partial^2 LL(\Phi)}{\partial b^2} \), \( DLL(\Phi) = \frac{\partial LL(\Phi)}{\partial b} \) and \( \Phi' = (b_{01}, b_{11}, b_{21}, b_{02}, b_{12}, b_{22}) \).

The Newton-Raphson method also incorporates the estimate of variance-covariance matrix of the maximum likelihood parameters \( \hat{\Phi} \), which is \( \{-[D^2 LL(b)]^{-1}\} \). A program in SAS version 5.18 PROC MATRIX [Sall (1982)] is written to obtain the solutions of the model. This program along with the complete derivations of the log-likelihood are provided in Appendix A.
4.3. Results

4.3.1. Recurrence-Free Survival

The Newton-Raphson method converges at fewer than 10 iterations with the values of the elements in DLL(\( \hat{b} \)) < 1x10^{-14}. The maximum likelihood estimates of the covariates coefficients and their standard deviations from the mixture model are presented in the following table.

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Group 1</th>
<th>Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant (( \hat{b}_0 ))</td>
<td>-2.95 (0.35)</td>
<td>-2.03 (0.23)</td>
</tr>
<tr>
<td>ER (( \hat{b}_1 ))</td>
<td>-0.12 (0.30)</td>
<td>-0.26 (0.20)</td>
</tr>
<tr>
<td>Tumor Diameter (( \hat{b}_2 ))</td>
<td>0.03 (0.06)</td>
<td>0.08 (0.03)</td>
</tr>
</tbody>
</table>

Using the covariate coefficients in Table 4.4, the hazard rates of the patients in group 1 and group 2 are calculated from equations (4.2.2) and (4.2.3), and their distributions are plotted in Figure 4.1. Note that the hazard rates in group 2 are well separated from the hazard rates in group 1. The hazard rates in group 2 are also distributed more widely than the hazard rates in group 1.

The hazard rates in groups 1 and 2 obtained from the mixture model without the covariates described in section 3.3 are presented in Table 4.5 along with their standard deviations. For comparison, the mean and the hazard rates obtained from the mixture model with the covariates are also presented in Table 4.5.
Figure 4.1. Distribution of individual hazard rates
Table 4.5. Hazard rate (standard deviation) for recurrence in groups 1 and 2.

<table>
<thead>
<tr>
<th>Mixture Model</th>
<th>Group 1</th>
<th>Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without Covariates</td>
<td>0.0526 (0.0066)</td>
<td>0.1437 (0.0129)</td>
</tr>
<tr>
<td>With Covariates</td>
<td>0.0528 (0.0042)</td>
<td>0.1509 (0.0446)</td>
</tr>
</tbody>
</table>

Table 4.5 shows that the model without the covariates yields slightly lower hazard rates compared to the model with the covariates, though this difference is not very apparent in group 1. Though it appears that there is very little difference, the mixture model with the covariates still has an advantage over the model without the covariates, since it can lead to a classification rule for a new observation.

For graphical representations, the estimated hazard rate of each patient in the training sample is used to calculate the average survival rate in groups 1 and 2 at time 0 to 15 years in increments of one year. The mixture survival rate is then computed as:

\[
\hat{S}(t) = 0.47 \hat{S}_{1}(t) + 0.53 \hat{S}_{2}(t),
\]

where

\[
\hat{S}_{1}(t) = \frac{1}{146} \sum_{k=1}^{146} \hat{S}_{k1}(t) = \frac{1}{146} \sum_{k=1}^{146} \exp(-\lambda_{k1}t)
\]

and

\[
\hat{S}_{2}(t) = \frac{1}{165} \sum_{k=1}^{165} \hat{S}_{k2}(t) = \frac{1}{165} \sum_{k=1}^{165} \exp(-\lambda_{k2}t)
\]

for \(t=0,1,2,\ldots,15\). The actuarial survival rates are calculated from the life table method obtained by the procedure LIFETEST in SAS. The component and the mixture survival distributions obtained from the model with the covariates are plotted in Figure 4.2. The actuarial survival distributions for the mixture sample, groups 1 and 2 are plotted along with the survival distributions obtained from the mixture models with and
without the covariates in Figures 4.3, 4.4, and 4.5, respectively.

From Figure 4.3 one can see that between three to eight years after mastectomy, the mixture model with the covariates provides a better fit to the actuarial curve compared to the mixture model without the covariates. After 10 years from mastectomy, the actuarial curve starts to level-off, while both curves from the mixture models decrease with similar rates. These patterns arise because the graphical representations of the mixture models assume a continuing rate of recurrence and are in a sense a time extrapolation. They do not assume that anyone is censored at each time interval as the actuarial does, and therefore their curves continue to decrease with time. The hump in the first two years of the actuarial curve is observed because the life table method starts with individuals who do not have the disease anymore and are not likely to have a recurrence within a short period of time, whereas the mixture models assume a constant hazard beginning at time 0. Similar patterns are noticed for groups 1 and 2 as shown in Figures 4.4 and 4.5, respectively.

To study the performance of the classification procedures, 20 patients are randomly excluded from the analysis, so that the training sample consists of only 291 patients. Furthermore, to get more accurate results, the analysis is performed 100 times in which 20 different patients are excluded randomly in every run and are classified using the estimated mixture parameters from the training sample. The classification results are then compared to the actual group membership to get the classification error rate. The percent classification error rate is computed as the number of misclassified cases in 20 patients. The classification error rates from 100 runs are tabulated in Appendix B. The summary consisting of the overall mean error rate, the standard deviation, the minimum and the maximum error rates from 100 runs is presented in Table 4.5.
Figure 4.2. Component and mixture distributions
Figure 4.3. Recurrence-free survival - All Patients
Figure 4.4. Recurrence-free survival - Group 1
Figure 4.5. Recurrence-free survival - Group 2
Table 4.5. Percent classification error rates from six classification procedures with ER and tumor diameter as the covariates and recurrence as the end point.

<table>
<thead>
<tr>
<th>Classification Procedure</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Neighborhood</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>43</td>
<td>47</td>
<td>34</td>
<td>33</td>
<td>32</td>
<td>48</td>
</tr>
<tr>
<td>S.D.\textsuperscript{a}</td>
<td>12</td>
<td>10</td>
<td>11</td>
<td>10</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>Min\textsuperscript{b}</td>
<td>20</td>
<td>25</td>
<td>10</td>
<td>15</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>Max\textsuperscript{c}</td>
<td>75</td>
<td>70</td>
<td>60</td>
<td>65</td>
<td>60</td>
<td>80</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Standard Deviation  
\textsuperscript{b} Minimum  
\textsuperscript{c} Maximum

From Table 4.5 one can see that, on the average, classification procedure 5 has the smallest mean error rate of 32% and standard deviation of 10%, though this is not significantly smaller than the 33% error rate from procedure 4 ($p=0.2$). The error rate from classification procedure 5 is slightly smaller compared to the error rate from classification procedure 3 of 34% ($p=0.09$). Classification procedure 1 has a higher error rate compared to procedure 5 (43% vs. 32%, $p < 0.001$), but it is still a better procedure than the non-parametric neighborhood method (43% vs. 48%, $p=0.001$). The error rate from classification procedure 2 is not statistically smaller than the error rate from the neighborhood classification procedure (47% vs. 48%, $p=0.3$). Better results are obtained by classifying everybody in group 2 than using procedure 2 or the neighborhood method. The proposed classification procedure 5 produces significantly improved results by 16% ($p < 0.001$) than the non-parametric neighborhood method.
4.3.2. Breast Cancer Survival

The Newton-Raphson method also converges in fewer than 10 iterations with the values of the elements in DLL(\(\hat{\beta}\)) \(< 1 \times 10^{-14}\). The maximum likelihood estimates of the covariates coefficients and their standard deviations are presented in the following table.

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Group 1</th>
<th>Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant ((\hat{\beta}_0))</td>
<td>-3.30 (0.38)</td>
<td>-2.21 (0.23)</td>
</tr>
<tr>
<td>ER ((\hat{\beta}_1))</td>
<td>-0.13 (0.33)</td>
<td>-0.48 (0.21)</td>
</tr>
<tr>
<td>Tumor Diameter ((\hat{\beta}_2))</td>
<td>0.04 (0.07)</td>
<td>0.06 (0.03)</td>
</tr>
</tbody>
</table>

From the estimated covariate coefficients in Table 4.6, the hazard rates of the training sample are calculated using equations (4.2.2) and (4.2.3), and plotted in Figure 4.6. Note that the hazard rates in group 1 are also well separated from the hazard rates in group 2. The distribution of the hazard rates in group 1 is not as widely spread as the distribution of the hazard rates in group 2. The mean and the standard deviation of the hazard rates in groups 1 and 2 obtained from the mixture model without the covariates are presented in Table 4.7, along with the mean hazard rates obtained from the mixture model with the covariates.
Figure 4.6. Distribution of individual hazard rates
Table 4.7. Hazard rate (standard deviation) for breast cancer survival in groups 1 and 2.

<table>
<thead>
<tr>
<th>Mixture Model</th>
<th>Group 1</th>
<th>Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without Covariates</td>
<td>0.0376 (0.0051)</td>
<td>0.0981 (0.0092)</td>
</tr>
<tr>
<td>With Covariates</td>
<td>0.0377 (0.0035)</td>
<td>0.1037 (0.0356)</td>
</tr>
</tbody>
</table>

Table 4.7 shows that the model without the covariates yields slightly lower hazard rates compared to the model with the covariates, though this difference is not very apparent for group 1. The model with the covariates still has an advantage over the model without the covariates, since it can lead to a classification rule for a new observation. The component and the mixture breast cancer survival distributions obtained from the model with the covariates are plotted in Figure 4.7. The actuarial survival distributions for all patients, groups 1 and 2 are plotted along with the survival distributions obtained from the mixture models with and without the covariates in Figures 4.8, 4.9, and 4.10, respectively.

Figure 4.8 shows that both mixture models with and without the covariates provide a very good fit to the actuarial curve, except in the first three years after mastectomy. The hump in the actuarial curve is observed because the life table method starts with individuals all of whom have early breast cancer and are not likely to die of the disease within a short period of time, whereas the mixture models again assume a constant hazard beginning at time 0. The fits for both of the mixture models for groups 1 and 2 are also quite good, as shown in Figures 4.9 and 4.10, respectively.
Figure 4.7. Component and mixture distributions
Figure 4.8. Breast cancer survival - All patients
Figure 4.9. Breast cancer survival - Group 1
Figure 4.10. Breast cancer survival - Group 2
To study the performance of the classification procedures when death from breast cancer is considered as the end point of survival, 20 patients are also randomly excluded from the training sample and the analysis is performed 100 times to get more accurate results. The classification error rates from five proposed classification procedures and the non-parametric neighborhood procedure obtained from 100 runs are tabulated in Appendix C. The overall mean, the standard deviation, the minimum and the maximum error rate from each classification procedure are presented in Table 4.8.

Table 4.8. Percent classification error rates from six classification procedures with ER and tumor diameter as the covariates and death from breast cancer as the end point.

<table>
<thead>
<tr>
<th>Classification Procedure</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Neighborhood</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>42</td>
<td>48</td>
<td>36</td>
<td>34</td>
<td>34</td>
<td>49</td>
</tr>
<tr>
<td>S.D.(^a)</td>
<td>11</td>
<td>12</td>
<td>12</td>
<td>10</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>Min(^b)</td>
<td>25</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>25</td>
</tr>
<tr>
<td>Max(^c)</td>
<td>75</td>
<td>70</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>70</td>
</tr>
</tbody>
</table>

\(^a\) Standard Deviation  
\(^b\) Minimum  
\(^c\) Maximum

From Table 4.8 one can see similar results with Table 4.5 when recurrence is considered the end point of the survival. Classification procedures 3, 4 and 5 seem to be superior than the rest of the classification procedures with the error rates of 34%, 34% and 36%, respectively. In fact the error rate from classification procedure 4 or 5 is not statistically smaller than the error rate from classification procedure 3 (p=0.1).
Classification procedure 1 produces the overall error rate of 42%, which is significantly greater than the error rate from procedure 5 ($p<0.001$). Classification procedure 2 and the non-parametric neighborhood method are shown again to have the worst misclassification rates of 48% and 49%, respectively. These two error rates are not significantly different from each other ($p=0.6$). Procedures 4 and 5 yield a 15% significant improvement in the classification error rate over the neighborhood procedure ($p < 0.001$), while procedure 3 has improved significantly by 13% over the neighborhood procedure ($p < 0.001$).

4.4. Discussion

Compared to the mixture model without the covariates, the model with the covariates produces a better fit to the actuarial curve between two and eight years after mastectomy for recurrence-free survival. The hump in the first two years of the actuarial curve is observed because the life table method starts with individuals who do not have the disease anymore and are not likely to have a recurrence within a short period of time, whereas the mixture models assume a constant hazard beginning at time 0. The actuarial curve starts to level-off at 10 years after mastectomy, while both curves from the mixture models decrease with similar rates, as shown in Figures 4.3, 4.4 and 4.5. These patterns are observed because the graphical representations of the mixture models assume a continuing rate of recurrence and are in a sense a time extrapolation. They do not assume that anyone is censored at each time interval as the actuarial does, and therefore the mixture model curves continue to decrease with time. For breast cancer survival, the curves from both models are almost overlapping with the actuarial curve, except in the first three years. The hump in the actuarial curve is also observed because the life table method starts with individuals all of whom have early breast
cancer and are not likely to die of the disease within a short period of time, whereas the mixture models again assume a constant hazard beginning at time 0.

The difference between the curves from the mixture models is more apparent for recurrence-free survival than breast cancer survival. This may be explained by the fact that the difference between the average hazard rates obtained from the mixture models with and without the covariates is slightly larger for recurrence-free than breast cancer survival, as shown in Tables 4.5 and 4.7. The excellent fit of both mixture distributions to the actuarial distribution for breast cancer survival may be due to weak relationships of the covariates on survival. It may be that the covariates are much more important for recurrence-free survival than for breast cancer survival.

One important improvement in using covariates with the mixture model is the capability of classifying new observations into one of the component populations based on these covariates. The classification procedures that use the survival times as well as the covariates produce more accurate results than the neighborhood procedure which uses the covariates only. Classification procedures 3, 4 and 5 produce better results than procedures 1 and 2 because the former use the mixture hazard rates, while the latter use the regular hazard rates. Though the smallest classification error rate is quite high (above 30%), this is still good considering that the correlations of the covariates with the grouping variable are weak, as shown in Table 4.9.

<table>
<thead>
<tr>
<th>Variable</th>
<th>PN</th>
<th>ER</th>
<th>Tumor Diameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>PN</td>
<td>1</td>
<td>-0.05</td>
<td>0.15</td>
</tr>
<tr>
<td>ER</td>
<td>-0.05</td>
<td>1</td>
<td>-0.04</td>
</tr>
<tr>
<td>Tumor Diameter</td>
<td>0.15</td>
<td>-0.04</td>
<td>1</td>
</tr>
</tbody>
</table>
To examine the performance of the classification procedures when different correlation sizes between the grouping variable and the covariates exist, a simulation study is conducted and the results are described in the next chapter.
CHAPTER 5
SIMULATION STUDY

5.1. Description of the Sample

In this chapter, a simulation study is conducted to investigate the behavior of the classification procedures for different types of samples. The sample is considered as a mixture of two components, i.e. groups 1 and 2, where group 1 reflects a better prognosis for survival than group 2. The mixing proportions \( \pi_1 \) and \( \pi_2 \) are known with \( \pi_1 \leq \pi_2 \), and \( \pi_2 = 1 - \pi_1 \). If the total sample size from the two groups is denoted by \( M \), then the number of observations in group 1 is \( n_1 = \pi_1 M \), and in group 2 is \( n_2 = \pi_2 M \). In a typical prospective clinical trial, the accrual period usually lasts for 5 years after the initiation of the trial. The proportion of patients who are censored will vary according to the follow-up period after accrual ends. The accrual period, \( A \), in this simulation study is set to 5 years. The time of the analysis, \( TL \), which reflects the follow-up period will be varied to obtain different proportions of censoring, \( c \). The first covariate, \( Z_1 \), is chosen to be of categorical type (0 or 1), with a value of one representing a better prognosis for the survival, and a value of zero representing a worse prognosis for the survival. The second covariate, \( Z_2 \), is of continuous type with greater values reflecting a worse prognosis for the survival. In addition, survival time, \( T \), will be generated from an exponential distribution. If the sum of \( A \) and \( T \) is greater than \( TL \), then the survival time is considered to be censored at \( TL - A \); otherwise it is considered as non-censored or an event with a value \( T \).

The grouping variable and the covariates are generated so that the correlations among them can be varied, in order to study the performance of the classification procedures under different types of conditions. Conditional distributions
are used to generate correlated variables [Law and Kelton (1982)]. Since the first
covariate, $Z_1$, is of the categorical type, the Bernoulli distribution will be used to
generate the values in each group. The second covariate, $Z_2$, which is of the continuous
type is assumed to be normally distributed in each group. The procedures to generate
correlated random variables can be given in three steps. First, generate the first
covariate, $Z_1$ from a Bernoulli distribution, so that the probability of obtaining a value
of 1 in group 1 ($s_1$) is higher than the probability of obtaining a value of 1 in group 2
($s_2$). Let $Z_{11}$ and $Z_{12}$ indicate these covariates in groups 1 and 2, respectively. Second,
generate the second covariate in group 1, $Z_{21}$, from a normal distribution such that $Z_{21}$
\[ \sim \text{Normal (} \mu_1, \sigma_1^2) \]. In the same manner, generate the second covariate in group 2,$Z_{22}$, from a normal distribution such that $Z_{22}$ \[ \sim \text{Normal (} \mu_2, \sigma_2^2) \] with $\mu_2 > \mu_1$ and
\[ \sigma_2^2 > \sigma_1^2. \] Third, generate the exponential survival time so that the survival
distributions in groups 1 and 2 are $S_1(t)$ and $S_2(t)$, respectively, where
\[ S_1(t) = \exp(-\lambda_1 t) = \exp\{-[\exp(a_{01}+a_{11}Z_1+a_{21}Z_2)] t\} \] and
\[ S_2(t) = \exp(-\lambda_2 t) = \exp\{-[\exp(a_{02}+a_{12}Z_1+a_{22}Z_2)] t\}. \] The coefficients $a_{01}$, $a_{11}$, $a_{21}$, $a_{02}$, $a_{12}$, and $a_{22}$
are chosen arbitrarily so that there is a good separation between the two groups. In this
simulation study, these values are chosen based on the results of the breast cancer
example in the previous chapter, which are $a_{01} = -2.99$, $a_{11} = -0.12$, $s_{21} = 0.03$,
$a_{02} = -2.00$, $a_{12} = -0.25$ and $a_{21} = 0.08$. To achieve different strength of
intercorrelations within the variables, the values of $s_1$, $s_2$, $\mu_1$, $\sigma_1$, $\mu_2$ and $\sigma_2$ will be varied.

5.2. Simulated Sample

In this section, the simulation of the data and the random number generators
used are described with great detail. The programs are written in the Statistical
Analysis System (SAS) version 5.18 PROC MATRIX [Sall (1982)]. The random number generators are those implemented in the version 5.18 SAS language written by Wallace (1985). Accrual times \( A_{k_i} \), first covariate \( Z_{1k_i} \), second covariate \( Z_{2k_i} \), exponential survival time \( t_{k_i} \), where \( i = 1 \) or \( 2 \) to denote group membership are generated as follows.

Accrual times \( \{A_{k_i} \mid k_i = 1, \ldots, \pi_i M\} \) are generated uniformly on the interval 0 to 5 using the SAS UNIFORM random number generator. The first covariate values, \( \{Z_{1k_i} \mid k_i = 1, \ldots, \pi_i M\} \), are generated using SAS RANBIN with a sample size of 1, from a Bernoulli distribution with \( s_1 \) as the probability of obtaining \( Z_{1k_i} = 1 \), and \((1-s_1)\) as the probability of obtaining \( Z_{1k_i} = 0 \). The second covariate values, \( \{Z_{2k_i} \mid k_i = 1, \ldots, \pi_i M\} \), which are of continuous type are generated using SAS RANNOR from a normal distribution with parameters \( (\mu_1, \sigma_1^2) \). The exponential survival times, \( \{t_{k_i} \mid k_i = 1, \ldots, \pi_i M\} \) are generated through the inverse-transform relationship between the survival distribution \( S(t) \) and the hazard rate \( \lambda_i \) given by \( S_i(t_{k_i}) = \exp(-\lambda_i t_{k_i}) \), then \(-\ln S_i(t_{k_i}) = \lambda_i t_{k_i} \) and \( t_{k_i} = \frac{-\ln S_i(t_{k_i})}{\lambda_i} \), where \( \lambda_i = a_0 + a_1 Z_1 + a_2 Z_2 \). The survival function, \( S_i(t_{k_i}) \) is generated so that it is Uniform [0,1] by using the SAS UNIFORM. For the \( k_i \)th observation, the sum of \( A_{k_i} \) and \( t_{k_i} \) is obtained, and then compared to the follow-up period, TL. If \( A_{k_i} + t_{k_i} \leq TL \), then observation is not censored with survival time \( t_{k_i} \), and if \( A_{k_i} + t_{k_i} > TL \), then observation is censored at \( TL - A_{k_i} \). The censoring mechanism is assumed to be noninformative, that is the distribution of the censoring times does not depend on the parameters of interest and furthermore, the censoring times are independent of the failure times.

The two sets of generated data from groups 1 and 2 are then concatenated to form a sample of size \( M \). In this simulation study, \( M \) is chosen to be 350 which is a typical sample size in a phase III cancer clinical trial. This value of \( M \) is also chosen to
minimize the computer processing time. The values of \( \pi_1 \) considered in the simulation are 0.50, 0.40, and 0.30. The values of \( \pi_2 \) in each of those cases will be equal to \( 1 - \pi_1 \). The times of analysis, \( T_A \), are chosen to be 10, 15 and 20 years to produce different proportions of censoring, \( c \), that are typical in node positive breast cancer clinical trial. Fifty observations are randomly excluded from the analysis and then classified based on their survival times and covariates.

5.3. Maximum Likelihood Estimates

From equation (3.2.2) in section 3.2, the log-likelihood function of the sample can be constructed as:

\[
LL(\Phi) = \sum_{k=1}^{M_1} \log \left[ p_1 \lambda_{1k} \exp \left\{ -\lambda_{1k} t_k \right\} \right] + \sum_{k=1}^{M_2} \log \left[ p_2 \lambda_{2k} \exp \left\{ -\lambda_{2k} t_k \right\} \right] + \\
\sum_{k=1}^{M_3} \log \left[ p_3 \exp \left\{ -\lambda_{1k} t_k \right\} \right] + \sum_{k=1}^{M_4} \log \left[ p_4 \exp \left\{ -\lambda_{2k} t_k \right\} \right],
\]

where \( M = M_1 + M_2 + M_3 + M_4 \); \( p_1 = \frac{M_1}{M} \) and is the proportion of non-censored observations in group 1; \( p_2 = \frac{M_2}{M} \) and is the proportion of non-censored observations in group 2; \( p_3 = \frac{M_3}{M} \) and is the proportion of censored observations in group 1; \( p_4 = \frac{M_4}{M} \) and is the proportion of censored observations in group 2; and \( \lambda_i = \exp(b_{0i} + b_{1i}Z_1 + b_{2i}Z_2) \) for \( i = 1, 2 \).

The maximum likelihood estimates of the unknown parameters, \( \hat{\Phi} \), are the solution of the following system of equations:

\[
\frac{\partial LL(\Phi)}{\partial b_{ji}} = 0, \ j = 0, 1, 2; \ i = 1, 2.
\]
The Newton-Raphson iterative method is used to obtain the approximate solutions to the log-likelihood function as:

\[ \Phi^{(s+1)} = \Phi^{(s)} - [D^2 LL(\Phi^{(s)})]^{-1} DLL(\Phi^{(s)}) \], \quad s = 0, 1, ... \]

where \( D^2 LL(\Phi) = \frac{\partial^2 LL(\Phi)}{\partial b^2} \), \( DLL(\Phi) = \frac{\partial LL(\Phi)}{\partial b} \) and \( \Phi' = (b_{01}, b_{11}, b_{21}, b_{02}, b_{12}, b_{22}) \). The program along with the complete derivations of the log-likelihood are written in SAS version 5.18 PROC MATRIX and they are available in Appendix D.

5.4. Results

One hundred samples are simulated at different combinations of the parameters \( \{s_1, s_2, \mu_1, \sigma_1, \mu_2, \sigma_2\} \) to obtain three sizes of correlations between the grouping variables and the two covariates, \( Z_1 \) and \( Z_2 \). These three sizes of correlations are shown in Table 5.1 and the different values of parameters to obtain these correlations are presented in Table 5.2.

<table>
<thead>
<tr>
<th>Type</th>
<th>Group vs. ( Z_1 )</th>
<th>Group vs. ( Z_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small</td>
<td>-0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Moderate</td>
<td>-0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Large</td>
<td>-0.8</td>
<td>0.8</td>
</tr>
</tbody>
</table>
Table 5.2. Values of parameters for different correlation sizes.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Small</th>
<th>Moderate</th>
<th>Large</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_1$</td>
<td>0.80</td>
<td>0.85</td>
<td>0.95</td>
</tr>
<tr>
<td>$s_2$</td>
<td>0.70</td>
<td>0.35</td>
<td>0.15</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>3.20</td>
<td>3.00</td>
<td>3.00</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>1.90</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>3.90</td>
<td>4.90</td>
<td>7.00</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>2.50</td>
<td>2.00</td>
<td>2.00</td>
</tr>
</tbody>
</table>

The program keeps track of the number of times the Newton-Raphson iteration method does not converge. A variable called Flag is introduced in the program to store the number of times the method does not converge. When the method does converge, the convergence is achieved within 10 iterations with the values of the elements in DLL(5) < 1x10^{-14}. The average classification error rates and their standard deviations from six classification procedures over 100 runs, along with the time of analysis (TL), the proportion of censoring (c), and the number of times the method does not converge (Flag) are presented in Tables 5.3, 5.4 and 5.5 for small, medium and high correlation, respectively.
Table 5.3. Average classification error rate (standard deviation) for small correlations between the grouping variable and the covariates, and when the proportion in group 1, $x_1=0.30$, 0.40, and 0.50.

<table>
<thead>
<tr>
<th>TLA</th>
<th>CB</th>
<th>Flag</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.44</td>
<td>0</td>
<td>49 (13)</td>
<td>33 (10)</td>
<td>34 (9)</td>
<td>34 (8)</td>
<td>36 (9)</td>
<td>30 (8)</td>
</tr>
<tr>
<td>15</td>
<td>0.27</td>
<td>0</td>
<td>51 (12)</td>
<td>33 (9)</td>
<td>29 (8)</td>
<td>26 (7)</td>
<td>28 (9)</td>
<td>31 (6)</td>
</tr>
<tr>
<td>20</td>
<td>0.18</td>
<td>0</td>
<td>47 (11)</td>
<td>31 (6)</td>
<td>25 (6)</td>
<td>23 (5)</td>
<td>26 (6)</td>
<td>30 (6)</td>
</tr>
</tbody>
</table>

$x_1 = 0.40$:

| 10  | 0.47| 0    | 47 (9)  | 41 (6)  | 34 (7) | 32 (7) | 33 (7) | 40 (6) |
| 15  | 0.31| 0    | 47 (9)  | 41 (6)  | 29 (8) | 27 (7) | 29 (8) | 41 (6) |
| 20  | 0.22| 0    | 44 (9)  | 40 (7)  | 29 (7) | 28 (6) | 29 (6) | 40 (7) |

$x_1 = 0.50$:

| 10  | 0.51| 0    | 43 (9)  | 49 (7)  | 38 (10) | 33 (7) | 34 (8) | 50 (7) |
| 15  | 0.35| 0    | 44 (7)  | 52 (7)  | 34 (9)  | 31 (8) | 32 (7) | 51 (6) |
| 20  | 0.25| 0    | 44 (7)  | 49 (7)  | 34 (7)  | 32 (7) | 32 (7) | 50 (7) |

a  Time of analysis  
b  Proportion of censored observations  
c  Number of times the Newton-Raphson method does not converge  
d  Neighborhood procedure
Table 5.4. Average classification error rate (standard deviation) for moderate correlations between the grouping variable and the covariates, and when the proportion in group 1, $\pi_1=0.30$, 0.40, and 0.50.

<table>
<thead>
<tr>
<th>TL</th>
<th>$c_b$</th>
<th>Flag</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>N_d</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.41</td>
<td>1</td>
<td>40 (21)</td>
<td>29 (6)</td>
<td>27 (9)</td>
<td>27 (7)</td>
<td>25 (8)</td>
<td>30 (6)</td>
</tr>
<tr>
<td>15</td>
<td>0.25</td>
<td>0</td>
<td>39 (20)</td>
<td>28 (6)</td>
<td>23 (7)</td>
<td>23 (6)</td>
<td>23 (7)</td>
<td>29 (6)</td>
</tr>
<tr>
<td>20</td>
<td>0.16</td>
<td>0</td>
<td>35 (19)</td>
<td>30 (7)</td>
<td>21 (7)</td>
<td>21 (6)</td>
<td>21 (6)</td>
<td>30 (7)</td>
</tr>
</tbody>
</table>

$\pi_1 = 0.40$:

<table>
<thead>
<tr>
<th>TL</th>
<th>$c_b$</th>
<th>Flag</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>N_d</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.45</td>
<td>1</td>
<td>34 (17)</td>
<td>37 (9)</td>
<td>31 (11)</td>
<td>29 (7)</td>
<td>28 (8)</td>
<td>39 (7)</td>
</tr>
<tr>
<td>15</td>
<td>0.29</td>
<td>0</td>
<td>33 (17)</td>
<td>40 (8)</td>
<td>25 (8)</td>
<td>25 (6)</td>
<td>24 (6)</td>
<td>41 (7)</td>
</tr>
<tr>
<td>20</td>
<td>0.20</td>
<td>0</td>
<td>33 (16)</td>
<td>39 (8)</td>
<td>26 (7)</td>
<td>27 (5)</td>
<td>25 (7)</td>
<td>40 (7)</td>
</tr>
</tbody>
</table>

$\pi_1 = 0.50$:

<table>
<thead>
<tr>
<th>TL</th>
<th>$c_b$</th>
<th>Flag</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>N_d</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.49</td>
<td>1</td>
<td>31 (13)</td>
<td>48 (10)</td>
<td>32 (12)</td>
<td>30 (8)</td>
<td>29 (8)</td>
<td>48 (8)</td>
</tr>
<tr>
<td>15</td>
<td>0.33</td>
<td>0</td>
<td>28 (12)</td>
<td>49 (10)</td>
<td>28 (9)</td>
<td>30 (7)</td>
<td>27 (8)</td>
<td>51 (8)</td>
</tr>
<tr>
<td>20</td>
<td>0.24</td>
<td>0</td>
<td>29 (13)</td>
<td>48 (10)</td>
<td>29 (8)</td>
<td>31 (7)</td>
<td>27 (8)</td>
<td>48 (8)</td>
</tr>
</tbody>
</table>

---

a Time of analysis
b Proportion of censored observations
c Number of times the Newton-Raphson method does not converge
d Neighborhood procedure
Table 5.5. Average classification error rate (standard deviation) for large correlations between the grouping variable and the covariates, and when the proportion in group 1, $\pi_1=0.30$, 0.40, and 0.50.

<table>
<thead>
<tr>
<th>TL$^a$</th>
<th>$c^b$</th>
<th>Flag$^c$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>N$^d$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>10</td>
<td>19</td>
<td>30 (28)</td>
<td>19 (11)</td>
<td>18 (10)</td>
<td>23 (7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>15</td>
<td>14</td>
<td>28 (27)</td>
<td>20 (10)</td>
<td>16 (9)</td>
<td>18 (5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>20</td>
<td>5</td>
<td>29 (28)</td>
<td>21 (10)</td>
<td>18 (8)</td>
<td>20 (6)</td>
</tr>
</tbody>
</table>

| $\pi_1 = 0.40$: |       |          | 10  | 11  | 19 (21) | 25 (14) | 19 (12) | 24 (6) | 18 (8) | 18 (10) |
|        |       |          | 15  | 12  | 24 (24) | 24 (14) | 20 (9)  | 23 (6) | 18 (7) | 19 (10) |
|        |       |          | 20  | 2   | 25 (23) | 27 (12) | 20 (10) | 26 (8) | 18 (9) | 21 (11) |

| $\pi_1 = 0.50$: |       |          | 10  | 8   | 22 (20) | 30 (18) | 23 (15) | 25 (6) | 21 (8) | 22 (12) |
|        |       |          | 15  | 4   | 26 (23) | 33 (16) | 23 (11) | 27 (7) | 19 (9) | 24 (12) |
|        |       |          | 20  | 0   | 21 (21) | 29 (16) | 23 (11) | 31 (8) | 17 (9) | 22 (13) |

$^a$ Time of analysis  
$^b$ Proportion of censored observations  
$^c$ Number of times the Newton-Raphson method does not converge  
$^d$ Neighborhood procedure
5.5. Discussion

From Tables 5.3, 5.4 and 5.5, one can see that the error rates from all of the classification procedures are affected by the size of the correlation. The error rates become smaller as the correlation size increases. Furthermore, the performance of the proposed classification procedures are also affected by the proportion of censoring in the data. On the average, their error rates get smaller as the proportion of censoring decreases. The neighborhood procedure, as expected, is not affected by the proportion of censoring since it does not use the survival times but only uses the covariates.

For small correlations between the grouping variable and the covariates, as shown in Table 5.3, all of the classification procedures do not work well when $\pi_1=0.30$, since their error rates are very close to the nominal error rate of 30%. Better results will be obtained by classifying everybody in group 2 than using any of the procedures, since the error rates will then be at the nominal value of $\pi_1$. When $\pi_1 = 0.40$ and 0.50, procedures 4 and 5 yield results which provide a greater improvement over the nominal rates than when $\pi_1=0.30$. The results from the other classification procedures do not improve as the value of $\pi_1$ increases. The best method for small correlation is classification procedure 4. The worst results are obtained from procedure 1 when $\pi_1=0.30$ or 0.40, or procedure 2 and the neighborhood method when $\pi_1=0.50$.

For moderate correlations between the grouping variable and the covariates, as shown in Table 5.4, all of the procedures still do not work very well for $\pi_1=0.30$, since the smallest error rate is above 20%. When $\pi_1=0.40$ and 0.50, procedure 5 is the best procedure for this type of correlation. The error rates from procedures 3 and 4 are very close to each other, making either one of them a second best classification procedure. Procedure 1 has better results compared to when the correlation size is small. The error rates from procedure 2 and the neighborhood method are the worst
and very close to the values of $\pi_I$.

Table 5.5 shows that for large correlations between the grouping variable and the covariates, procedure 5 is still superior than the rest of the classification procedures. The neighborhood classification procedure has improved significantly since its error rate becomes considerably smaller as the correlation size increases. Procedures 3 and 4 yield comparable results to the neighborhood method. The error rates from procedures 1 and 2 are smaller than those when the correlation size is moderate, though they are still the highest compared to those from the other procedures when the correlation size is large.

As a general guideline, when small or moderate correlations between the grouping variable and the covariates are present, classification procedures 4 and 5 are the best methods to use. Procedure 5 is still the best method to use for large correlations, followed by the neighborhood procedure. In conclusion, the classification procedures that use the mixture hazard rate instead of the regular hazard rate produce more accurate results as demonstrated by the performances of procedures 3, 4 or 5 versus 1 or 2. This simulation study also shows that even when the correlation size is large, then classification procedure 5 produces better results than the neighborhood method, while the other proposed classification procedures yield comparable results to the neighborhood method. Since highly correlated data are rarely encountered in practice, procedure 5 which uses both survival times and covariates is a better alternative to the neighborhood method which uses covariates alone, or to procedures 1 and 2 which use the regular hazard rates. Furthermore, when the covariates are highly correlated with the grouping variable, the classification procedures are not as compelling and of interest as when the correlation is small.
CHAPTER 6
CONCLUSIONS AND FURTHER RESEARCH

6.1. Summary and Conclusions

The problem and the objectives are stated in chapter 1. The survival
distribution of observations from a heterogeneous population is considered to be a
mixture of two component distributions, and the incorporation of covariates into this
finite mixture survival distribution is proposed. The concept of traditional finite
mixture distributions and classical discriminant analyses are reviewed and compared
with the concept of the proposed research method. The primary objective of this
research is to incorporate covariates in finite mixture survival distributions so that
classification of new observations into one of the component populations can be
performed based on their survival times and covariates.

Chapter 2 consists of the literature review on finite mixture distributions and
covariates, and includes the definition of finite mixture distributions, the identifiability,
the sampling types and the methods of estimation in finite mixture distributions. In
addition, multivariate statistical techniques that have been used for analyzing survival
data with covariates are reviewed. The censoring mechanism used in this research is
also described in this chapter.

The mixture model with covariates is postulated in chapter 3. The survival
distribution of observations in the study population is considered to be a mixture of two
exponential distributions. Covariates are incorporated in the mixture survival
distributions by expressing the hazard rate of each component distribution in the
training sample as a log-linear function of the covariates. The log-likelihood function
of the sample is constructed, and the iterative procedure using the Newton-Raphson
method to solve for the unknown parameters is described. The mixture model with covariates is compared to the mixture model without covariates and the life table method. Five procedures for classifying new observations with certain covariates into one of the component populations are proposed. These classification procedures are compared to the neighborhood non-parametric procedure.

The models are illustrated using data from a prospective adjuvant clinical trial in breast cancer based at Case Western Reserve University in chapter 4. The patients are considered to be a mixture of those with axillary positive nodes (PN) between 1 and 3 (group 1) and those with PN greater than four (group 2). In this example, analyses using recurrence and death from breast cancer as the end points are considered. Two covariates, estrogen receptor status and tumor diameter, are used in the mixture model, because these two covariates are known to be important prognostic factors in breast cancer. The results from the model comparisons indicate that there is very little difference between the mixture models with and without the covariates. However, an advantage is gained by using the covariates in the mixture model, that is the capability to classify observations with unknown group memberships based on their survival times and covariates. Three of the five proposed classification procedures produce better results compared to the neighborhood procedure.

To examine the performances of the classification procedures under different types of samples, a simulation study is conducted in chapter 5. Three sizes of correlations between the grouping variable and the covariates are generated, which are small, moderate and large. The results from the simulation study suggest that when small or moderate correlations between the grouping variable and the covariates are present, classification procedures 3, 4 and 5 that use both the survival times and covariates give more accurate results than procedures 1 and 2 and the neighborhood
procedure that uses the covariates alone. When the correlations between the grouping variable and the covariates are large, classification procedure 5 yields the best results, while the other procedures yield comparable results to the neighborhood method. Highly correlated data are rarely encountered in practice, however the use of covariates with finite mixture distributions provides a better classification method than the neighborhood method even in this case.

6.2. Suggestions for Further Research

The use of other parametric distributions such as the Weibull or the Gompertz can be explored. One disadvantage of using a distribution other than the exponential is that there are more unknown parameters to estimate. The possible models for the hazard function, h, for the Weibull and Gompertz distributions are, respectively:

\[
a) \quad h(t; \alpha, \delta) = \alpha \delta t^{\delta - 1} = g(Z) \delta t^{\delta - 1} \\
\]

and

\[
b) \quad h(t; \alpha, \delta) = e^{\alpha + \delta t} = e^{g(Z) + \delta t} ,
\]

where \( g(Z) = \exp(\beta'Z) \).

In this research, the log-linear model is used to express the relationship between the covariates and the hazard function. One may use the linear exponential function by Feigl and Zelen (1965), or by Bryar et al. (1974) to model the covariates. These models are, respectively:

\[
\mu = \frac{1}{\lambda} = \sum_{j=0}^{u} b_j Z_j
\]

and

\[
\lambda = \sum_{j=0}^{u} b_j Z_j
\]
In the simulation study, a log-normal distribution can be used instead of the normal distribution to generate the second covariate values. More covariates can also be used in the simulation study to examine whether the performances of the classification procedures will improve as more covariates are used in the model. The concept of incorporating covariates in finite mixture survival distributions is still new. More simulation studies as well as applications need to be performed to get a better understanding of the model.
APPENDIX A
SAS PROGRAM FOR APPLICATION IN BREAST CANCER

******************************************************************************:
PROC MATRIX;

***THIS IS TO LOAD THE DATA BASE AND IDENTIFY EACH VARIABLE***:
   FETCH A2 DATA=BRAWCOCT91B;
   PN = A2(4);
   ER = A2(3); * THIS IS ER COVARIATE;
   TD = A2(1); * THIS IS TD COVARIATE;
 * T = A2(9); * THIS IS FOR RECURRENT;
   T = A2(10); * THIS IS FOR DEATH FROM BREAST CANCER;
   CR = A2(6); * THIS IS FOR RECURRENT;
   CR = A2(7); * THIS IS FOR DEATH FROM BREAST CANCER:

******************************************************************************:

*** CONCATENATE THE VARIABLES TO FORM MATRIX DAT***:
* AND INTRODUCE THE INDICATORS VARIABLES C1 - C4*:
   DAT = PN || ER || TD || T || CR;
   NTOT = NROW(DAT);
   IND = DAT(5);
   C1 = J(NTOT,1.0);
   C2 = J(NTOT,1.0);
   C3 = J(NTOT,1.0);
   C4 = J(NTOT,1.0);
   PN = DAT(1);
   DO K=1 TO NTOT;
      IF PN(K) = 1 AND IND(K) = 1 THEN C1(K) = 1;
      ELSE C1(K) = 0;
      IF PN(K) = 2 AND IND(K) = 1 THEN C2(K) = 1;
      ELSE C2(K) = 0;
      IF PN(K) = 1 AND IND(K) = 0 THEN C3(K) = 1;
      ELSE C3(K) = 0;
      IF PN(K) = 2 AND IND(K) = 0 THEN C4(K) = 1;
      ELSE C4(K) = 0;
      END;
   INP = DAT || C1 || C2 || C3 || C4;
   NOSIM=100;
   FLAG=0;
CENS = J(NOSIM,1,0);
TERR = J(NOSIM,6,0);
TERR_2 = J(NOSIM,6,0);
FNPAR = J(NOSIM,6,0);
FVAR = J(NOSIM,6,0);
COR12 = J(NOSIM,1,0);
COR13 = J(NOSIM,1,0);
COR14 = J(NOSIM,1,0);
COR23 = J(NOSIM,1,0);
COR24 = J(NOSIM,1,0);
COR34 = J(NOSIM,1,0);

DO SIM=1 TO NOSIM;

******THIS IS TO EXCLUDE 20 OBSERVATIONS RANDOMLY*******:

    U=INT(UNIFORM(J(20,1,0))#NTOT)+1;
    ALL = (1:NTOT)';
    ALL(U,) = 0;
    NINP = INP(LOC(ALL),);
    REST = INP(U,);
    *PRINT NINP REST;

*******************************************************************:

******THIS IS TO COMPUTE THE CORRELATION COEFFICIENTS*******:

    X = NINP(1) || NINP(2) || NINP(3) || NINP(5);
    NUMER = NROW(X);
    X_SUM = X(+,);
    XPX = X'X-X_SUM'X_SUM#/NUMER;
    SCOR = 1#/SQRT(DIAG(XPX));
    CORR = SCOR*XPX*SCOR;
    * PRINT CORR;
    COR12(SIM,) = CORR(1,2);
    COR13(SIM,) = CORR(1,3);
    COR14(SIM,) = CORR(1,4);
    COR23(SIM,) = CORR(2,3);
    COR24(SIM,) = CORR(2,4);
    COR34(SIM,) = CORR(3,4);

*******************************************************************:
***THIS IS TO IDENTIFY THE TRAINING SAMPLE***;
NINP1 = NINP(LOC(NINP(.1)=1),);
NINP2 = NINP(LOC(NINP(.1)=2),);
*NPRINT NINP NINP2 NINP1;
N1 = NROW(NINP1);
N2 = NROW(NINP2);
N=N1+N2;
*NPRINT NTOT N N1 N2;

******************************************************************************;

***THIS IS TO COMPUTE THE INITIAL ESTIMATES THROUGH LEAST***:

******************************************************************************;

R1 = J(N1,1,0);
R2 = J(N2,1,0);
S1_EST = J(N1,1,0);
S2_EST = J(N2,1,0);
R1 = RANK(NINP1(.4));
R2 = RANK(NINP2(.4));
IND1 = NINP1(.5);
IND2 = NINP2(.5);
T1 = NINP1(.4);
T2 = NINP2(.4);
PN1 = NINP1(.1);
PN2 = NINP2(.1);
ER1 = NINP1(.2);
ER2 = NINP2(.2);
TD1 = NINP1(.3);
TD2 = NINP2(.3);
L1 = J(N1,1,0);
L2 = J(N2,1,0);
X01 = J(N1,1,1);
X02 = J(N2,1,1);
PAR1=J(3,1,0);
PAR2=J(3,1,0);

DO M=1 TO N1:
IF IND1(M.) = 1 THEN S1_EST(M.) = (N1-R1(M.))/((N1+1));
ELSE S1_EST(M.) = 0;
IF S1_EST(M.) = 0 THEN L1(M.) = LOG(-LOG(S1_EST(M.)) / T1(M.));
ELSE L1(M.) = 0;
IF L1(M.) = 0 THEN DO;
    ER1(M.) = 0;
    TD1(M.) = 0;
    X01(M.) = 0;
END;
END;
DO M=1 TO N2;
  IF IND2(M,.) = 1 THEN S2_EST(M,.) = (N2-R2(M,.)+1)#/(N2+1);
  ELSE S2_EST(M,.) = 0;
  IF S2_EST(M,.) = 0 THEN L2(M,.) = LOG(-LOG(S2_EST(M,.)) #/ T2(M,.));
  ELSE L2(M,.) = 0;
  IF L2(M,.) = 0 THEN DO;
    ER2(M,.) = 0;
    TD2(M,.) = 0;
    X02(M,.) = 0;
  END;
END;

TEMP1 = L1 || X01 || ER1 || TD1;
TEMP2 = L2 || X02 || ER2 || TD2;
NTEMP1 = TEMP1(LOC(TEMP1(:,1) = 0),);
NTEMP2 = TEMP2(LOC(TEMP2(:,1) = 0),);
PRINT TEMP1 NTEMP1;

X1= NTEMP1(:,2) || NTEMP1(:,3) || NTEMP1(:,4);
X2= NTEMP2(:,2) || NTEMP2(:,3) || NTEMP2(:,4);
PAR1 = INV(X1'*X1)*X1'*NTEMP1(:,1);
PAR2 = INV(X2'*X2)*X2'*NTEMP2(:,1);
PAR = PAR1 // PAR2;

**********THIS IS THE NEWTON-RAPHSON PROCEDURE**********:;

NPAR = PAR;
Z1 = NINP(,2); Z2 = NINP(,3); T=NINP(,4);
C1 = NINP(,6); C2 = NINP(,7);
C3 = NINP(,8); C4 = NINP(,9);
CC1=SUM(C1); CC2=SUM(C2);
CC3=SUM(C3); CC4=SUM(C4);
N=CC1+CC2+CC3+CC4;
P1=CC1#/N; P2=CC2#/N;
P3=CC3#/N; P4=1-P1-P2-P3;
U=J(6,1,0);
I=J(6,6,0);
COUNT=0;

BEGIN: PAR=NPAR;
B01=PAR(1,);
B11=PAR(2,);
B21=PAR(3,);
B02=PAR(4,);
B12=PAR(5,);
B22=PAR(6,);

LL=SUM(C1*(LOG(P1)+B01+B11#Z1+B21#Z2T*(EXP(B01+B11#Z1+B21#Z2))))
  +C2*(LOG(P2)+B02+B12#Z1+B22#Z2T*(EXP(B02+B12#Z1+B22#Z2)))
  +C3*(LOG(P3)-T*(EXP(B01+B11#Z1+B21#Z2)))
  +C4*(LOG(P4)-T*(EXP(B02+B12#Z1+B22#Z2))));
******* THESE ARE THE FIRST DERIVATIVES OF THE LOG-LIKELIHOOD *******;

\[ U(1,) = \text{SUM}(C1\#(1-(\text{EXP}(B01+B11#Z1+B21#Z2))#T)-
C3\#((\text{EXP}(B01+B11#Z1+B21#Z2))#T)); \]

\[ U(2,) = \text{SUM}(C1\#(Z1#Z1#T#(\text{EXP}(B01+B11#Z1+B21#Z2)))-
C3\#(Z1#T#(\text{EXP}(B01+B11#Z1+B21#Z2)))); \]

\[ U(3,) = \text{SUM}(C1\#(Z2#Z2#T#(\text{EXP}(B01+B11#Z1+B21#Z2)))-
C3\#(Z2#T#(\text{EXP}(B01+B11#Z1+B21#Z2)))); \]

\[ U(4,) = \text{SUM}(C2\#(1-(\text{EXP}(B02+B12#Z1+B22#Z2))#T)-
C4\#((\text{EXP}(B02+B12#Z1+B22#Z2))#T)); \]

\[ U(5,) = \text{SUM}(C2\#(Z1-Z1#T#(\text{EXP}(B02+B12#Z1+B22#Z2)))-
C4\#(Z1#T#(\text{EXP}(B02+B12#Z1+B22#Z2)))); \]

\[ U(6,) = \text{SUM}(C2\#(Z2-Z2#T#(\text{EXP}(B02+B12#Z1+B22#Z2)))-
C4\#(Z2#T#(\text{EXP}(B02+B12#Z1+B22#Z2)))); \]

***************************************************************************************;

***** THESE ARE THE SECOND DERIVATIVES OF THE LOG-LIKELIHOOD *****;

\[ I(1,1) = \text{SUM}(-C1\#((\text{EXP}(B01+B11#Z1+B21#Z2))#T)-
C3\#((\text{EXP}(B01+B11#Z1+B21#Z2))#T)); \]

\[ I(1,2) = \text{SUM}(-C1\#(Z1#T#(\text{EXP}(B01+B11#Z1+B21#Z2)))-
C3\#(Z1#T#(\text{EXP}(B01+B11#Z1+B21#Z2)))); \]

\[ I(1,3) = \text{SUM}(-C1\#(Z2#T#(\text{EXP}(B01+B11#Z1+B21#Z2)))-
C3\#(Z2#T#(\text{EXP}(B01+B11#Z1+B21#Z2)))); \]

\[ I(1,4) = 0; \quad I(1,5) = 0; \quad I(1,6) = 0; \]

\[ I(2,1) = I(1,2); \]

\[ I(2,2) = \text{SUM}(-C1\#(Z1-Z1#T#(\text{EXP}(B01+B11#Z1+B21#Z2)))-
C3\#(Z1#Z1#T#(\text{EXP}(B01+B11#Z1+B21#Z2)))); \]

\[ I(2,3) = \text{SUM}(-C1\#(Z2-Z2#T#(\text{EXP}(B01+B11#Z1+B21#Z2)))-
C3\#(Z2#Z2#T#(\text{EXP}(B01+B11#Z1+B21#Z2)))); \]

\[ I(2,4) = 0; \quad I(2,5) = 0; \quad I(2,6) = 0; \]

\[ I(3,1) = I(1,3); \quad I(3,2) = I(2,3); \]

\[ I(3,3) = \text{SUM}(-C1\#(Z2#Z2#T#(\text{EXP}(B01+B11#Z1+B21#Z2)))-
C3\#(Z2#Z2#T#(\text{EXP}(B01+B11#Z1+B21#Z2)))); \]

\[ I(3,4) = 0; \quad I(3,5) = 0; \quad I(3,6) = 0; \]
I(4,1)=I(1,4); I(4,2)=I(2,4); I(4,3)=I(3,4);
I(4,4)=SUM(-C2#((EXP(B02+B12#Z1+B22#Z2))#T)-C4#((EXP(B02+B12#Z1+B22#Z2))#T));
I(4,5)=SUM(-C2#(Z1#T#((EXP(B02+B12#Z1+B22#Z2))^11)-C4#(Z1#T#((EXP(B02+B12#Z1+B22#Z2))^11));
I(4,6)=SUM(-C2#(Z2#T#((EXP(B02+B12#Z1+B22#Z2))^11)-C4#(Z2#T#((EXP(B02+B12#Z1+B22#Z2))^11));
I(5,1)=I(1,5); I(5,2)=I(2,5);
I(5,3)=I(3,5); I(5,4)=I(4,5);
I(5,5)=SUM(-C2#(Z1#Z1#T#(EXP(B02+B12#Z1+B22#Z2)))-C4#(Z1#Z1#T#(EXP(B02+B12#Z1+B22#Z2))));
I(5,6)=SUM(-C2#(Z1#Z2#T#(EXP(B02+B12#Z1+B22#Z2)))-C4#(Z1#Z2#T#(EXP(B02+B12#Z1+B22#Z2))));
I(6,1)=I(1,5); I(6,2)=I(2,6); I(6,3)=I(3,6);
I(6,4)=I(4,6); I(6,5)=I(5,6);
I(6,6)=SUM(-C2#(Z2#Z2#T#(EXP(B02+B12#Z1+B22#Z2)))-C4#(Z2#Z2#T#(EXP(B02+B12#Z1+B22#Z2))));

*THIS IS THE ITERATION STEPS*;
IF DET(I) = 0 THEN DO:
    FLAG = FLAG + 1;
    GOTO OUT;
END;
ELSE GOTO CONT;

CONT: INVI = INV(I);
NPAR = PAR - INVI*U;
COUNT = COUNT + 1;
IF COUNT < 10 THEN GOTO BEGIN;
ELSE GOTO FIN;

FIN: D=DET(INVI);
PRINT LL D I INVI NPAR;

FNPAR((,) = NPAR';
FVAR((,) = INVI(1,1) || INVI(2,2) || INVI(3,3) || INVI(4,4) || INVI(5,5) || INVI(6,6) ;
***************RECALCULATE HAZARD FOR EACH GROUP***************;
HZ1= (C1 # (EXP(NPAR(1,)+NPAR(2,)#Z1+NPAR(3,)#Z2))) + 
     (C3 # (EXP(NPAR(1,)+NPAR(2,)#Z1+NPAR(3,)#Z2))) ;
HZ2= (C2 # (EXP(NPAR(4,)+NPAR(5,)#Z1+NPAR(6,)#Z2))) + 
     (C4 # (EXP(NPAR(4,)+NPAR(5,)#Z1+NPAR(6,)#Z2))) ;
H1 = HZ1(LOC(HZ1>0,));
H2 = HZ2(LOC(HZ2>0,));
HH1 = H1; HH2=H2;
H1(RANK(H1,)) = HH1;
H2(RANK(H2,)) = HH2;
H = H1 // H2;
* PRINT H ;
MH1 = SUM(H1)#/N1;
MH2 = SUM(H2)#/N2;
S11 = (SUM((H1'-MH1)*(H1'-MH1)')) #/ (N1-1);
S22 = (SUM((H2'-MH2)*(H2'-MH2)')) #/ (N2-1);
SP = ((N1-1)*S11+(N2-1)*S22)/(N1+N2-2);
STDP= SQRT(SP);
* PRINT MH1 MH2 N1 N2 S11 S22 STDP;

PC1 = P1 + P3; *THIS IS PRIOR GROUP 1 MEMBERSHIP PROBABILITY:
PC2 = P2 + P4; *THIS IS PRIOR GROUP 2 MEMBERSHIP PROBABILITY:
CENS(SIM,.) = P3 + P4; *THIS IS PROPORTION CENSORED:

NEWNP = REST;
NOBS = NCROW(NEWNP);
NIND1 = J(NOBS,1,0);
NIND2 = J(NOBS,1,0);
NIND3 = J(NOBS,1,0);
NIND4 = J(NOBS,1,0);
ACT._CL = NEWNP(1);
NZ1 = NEWNP(2);
NZ2 = NEWNP(3);
NT  = NEWNP(4);
NIND = NEWNP(5);

DO I=1 TO NOBS;
   IF NIND(I,) = 1 THEN DO;
      NIND1(I,)= 1; NIND2(I,)= 1;
      NIND3(I,)= 0; NIND4(I,)= 0;
   END;

   IF NIND(I,)= 0 THEN DO;
      NIND1(I,)= 0; NIND2(I,)= 0;
      NIND3(I,)= 1; NIND4(I,)= 1;
   END;
END;
\[ \begin{align*}
\text{H1NEW} &= \text{EXP}(\text{NPAR}(1.)+(\text{NPAR}(2.)\#\text{NZ1})+(\text{NPAR}(3.)\#\text{NZ2})); \\
\text{H2NEW} &= \text{EXP}(\text{NPAR}(4.)+(\text{NPAR}(5.)\#\text{NZ1})+(\text{NPAR}(6.)\#\text{NZ2})); \\
\text{F1} &= \text{H1NEW} \# (\text{EXP}(\text{H1NEW}\#\text{NT})); \\
\text{F2} &= \text{H2NEW} \# (\text{EXP}(\text{H2NEW}\#\text{NT})); \\
\text{S1} &= \text{EXP}(\text{H1NEW}\#\text{NT}); \\
\text{S2} &= \text{EXP}(\text{H2NEW}\#\text{NT}); \\
\text{FMIX} &= (\text{P1}\#\text{F1})+(\text{P2}\#\text{F2}); \\
\text{SMIX} &= (\text{P3}\#\text{S1})+(\text{P4}\#\text{S2}); \\
\text{P1}_T &= ((\text{NIND1}\#(\text{P1} \# \text{F1}))+((\text{NIND3}\#(\text{P3} \# \text{S1}))) \#/
((\text{NIND1}\#(\text{FMIX}))+((\text{NIND3}\#(\text{SMIX}))) \\
\text{P2}_T &= ((\text{NIND2}\#(\text{P2} \# \text{F2}))+((\text{NIND4}\#(\text{P4} \# \text{S2}))) \#/
((\text{NIND2}\#(\text{FMIX}))+((\text{NIND4}\#(\text{SMIX}))) \\
\text{F}_T &= \text{P1}_T\#\text{F1} + \text{P2}_T\#\text{F2}; \\
\text{S}_T &= \text{P1}_T\#\text{S1} + \text{P2}_T\#\text{S2}; \\
\text{H}_T &= \text{F}_T \#/ \text{S}_T; \\
\end{align*} \]

***THIS PART IS FOR DIFFERENT CLASSIFICATION PROCEDURES***:

\[ \begin{align*}
\text{N1} &= \text{NROW(H1)}; \\
\text{N2} &= \text{NROW(H2)}; \\
\text{N\_NEW} &= \text{NROW(H\_T)}; \\
\text{CL1} &= \text{J(N\_NEW,1,0)}; \\
\text{CL2} &= \text{J(N\_NEW,1,0)}; \\
\text{CL3} &= \text{J(N\_NEW,1,0)}; \\
\text{CL4} &= \text{J(N\_NEW,1,0)}; \\
\text{CL5} &= \text{J(N\_NEW,1,0)}; \\
\text{CLN} &= \text{J(N\_NEW,1,0)}; \\
\text{ESTPR1} &= \text{J(N\_NEW,1,0)}; \\
\text{ESTPR2} &= \text{J(N\_NEW,1,0)}; \\
\text{PR1A} &= \text{J(N\_NEW,1,0)}; \\
\text{PR2A} &= \text{J(N\_NEW,1,0)}; \\
\text{PR3A} &= \text{J(N\_NEW,1,0)}; \\
\text{PR4A} &= \text{J(N\_NEW,1,0)}; \\
\text{ERR1} &= \text{J(N\_NEW,1,0)}; \\
\text{ERR2} &= \text{J(N\_NEW,1,0)}; \\
\text{ERR3} &= \text{J(N\_NEW,1,0)}; \\
\text{ERR4} &= \text{J(N\_NEW,1,0)}; \\
\text{ERR5} &= \text{J(N\_NEW,1,0)}; \\
\text{ERRN} &= \text{J(N\_NEW,1,0)}; \\
\end{align*} \]
**THIS IS CLASSIFICATION METHOD 1 (BASED ON 2 HAZARD VALUES);**

```
DO J = 1 TO N_NEW;
    K = 1;
    DO K = 1 TO N1;
        IF H1NEW(J) <= H1(K) THEN GOTO COUNT1;
        END;
    COUNT1: PR1A(J) = (N1-(K-1))#/N1;
    END;

DO J = 1 TO N_NEW;
    K = 1;
    DO K = 1 TO N2;
        IF H2NEW(J) < H2(K) THEN GOTO COUNT4;
        END;
    COUNT4: PR4A(J) = (K-1)#/N2;
    END;

DO L=1 TO N_NEW;
    IF PR1A(L) > PR4A(L) THEN CL1(L) = 1;
    IF PR1A(L) <= PR4A(L) THEN CL1(L) = 2;
    IF CL1(L)=ACT_CL(L) THEN ERR1(L)=0;
    ELSE ERR1(L) = 1;
    END;
```

* PR1 = NZ1 | | NZ2 | | PR1A | | PR4A | | CL1 | | ACT_CL:
  TERR1 = INT((SUM(ERR1)#/N_NEW)*100);
* PRINT PR1 TERR1;

**THIS IS FOR CLASSIFICATION METHOD 2 (BASED ON 2 HAZARD VALUES);
AND COMPARE THE BAYES POSTERIOR PROBABILITIES:**

```
DO J = 1 TO N_NEW;
    K = 1;
    DO K = 1 TO N1;
        IF H1NEW(J) <= H1(K) THEN GOTO COUNT2;
        END;
    COUNT2: PR2A(J) = (N1-(K-1))#/N1;
    END;

DO J = 1 TO N_NEW;
    K = 1;
    DO K = 1 TO N2;
        IF H1NEW(J) < H2(K) THEN GOTO COUNT3;
        END;
    COUNT3: PR3A(J) = (K-1)#/N2;
    END;
```
PROM1 = J(N_NEW,1,0);
PROM2 = J(N_NEW,1,0);
PRO1 = J(N_NEW,1,0);
PRO2 = J(N_NEW,1,0);

DO L=1 TO N_NEW;
PROM1(L) = PR1A(L) + PR3A(L) + PROM2(L) = PR2A(L) + PR4A(L) + PROM1(L);
PRO1(L) = PR1A(L) + PC1 + PROM1(L);
PRO2(L) = PR4A(L) + PC2 + PROM2(L);
IF PRO1(L) > PRO2(L) THEN CL2(L) = 1;
ELSE ERR2(L) = 0;
END;

* PR2 = NZ1 | | NZ2 | | H1NEW | | H2NEW | | PRO1 | | PRO2 | | CL2 | | ACT_CL:
   TERR2 = INT((SUM(ERR2)#/N_NEW)*100);
* PRINT PR2 TERR2;

******************************************************************************

******************************************************************************
* THIS IS CLASSIFICATION METHOD 3 (BASED ON MIXTURE HAZARD);
******************************************************************************

DO J = 1 TO N_NEW;
   K = 1;
   DO K = 1 TO N1;
       IF H_T(J) < H1(K) THEN GOTO OUT1;
   END;
   OUT1: ESTPR1(J) = (N1-(K-1))#/N1;

END:

DO J = 1 TO N_NEW;
   K = 1;
   DO K = 1 TO N2;
       IF H_T(J) < H2(K) THEN GOTO OUT2;
   END;
   OUT2: ESTPR2(J) = (K-1)#/N2;

END:

DO L=1 TO N_NEW;
   IF ESTPR1(L) > ESTPR2(L) THEN CL3(L) = 1;
   IF ESTPR1(L) <= ESTPR2(L) THEN CL3(L) = 2;
   IF CL3(L) = ACT_CL(L) THEN ERR2(L) = 0;
   ELSE ERR3(L) = 1;
END;

* PR3 = NZ1 | | NZ2 | | NT | | NIND | | H_T | | ESTPR1 | | ESTPR2 | | CL3 | | ACT_CL:
   TERR3 = INT((SUM(ERR3)#/N_NEW)*100);
* PRINT PR3 TERR3;

******************************************************************************
*THIS IS FOR THE CLASSIFICATION METHOD 4 (COMPARE POST. PROB.);*

```plaintext
DO J = 1 TO N_NEW;
    IF P1_T(J) > P2_T(J) THEN CL4(J) = 1;
    IF P1_T(J) <= P2_T(J) THEN CL4(J) = 2;
    IF CL4(J) = ACT_CL(J) THEN ERR4(J) = 0;
    ELSE ERR4(J) = 1;
END;
* PR4 = N21 | | N22 | | NT | | NIND | | P1_T | | P2_T | | CL4 | | ACT_CL;
* TERR4 = INT((SUM(ERR4)#/N_NEW)*100);
* PRINT PR4 TERR4;
```

*THIS IS CLASSIFICATION 5 ADAPTED FROM NEIGHBORHOOD METHOD**;

```plaintext
DIFF1 = J(N1,1,0); DIFF2 = J(N2,1,0);
DIST1 = J(N1,1,0); DIST2 = J(N2,1,0);
NDIST1 = J(N_NEW,1,0);
NDIST2 = J(N_NEW,1,0);

DO I = 1 TO N_NEW;
    DO K = 1 TO N1;
        DIFF1(K) = H1(K) - H_T(I);
        DIST1(K) = DIFF1(K)*INV(SP)*DIFF1(K);  
    END;
    NDIST1(I) = MIN(DIST1);
END;

DO I = 1 TO N_NEW;
    DO K = 1 TO N2;
        DIFF2(K) = H2(K) - H_T(I);
        DIST2(K) = DIFF2(K)*INV(SP)*DIFF2(K);  
    END;
    NDIST2(I) = MIN(DIST2);
END:

DIST = NDIST1 | | NDIST2;
DO I = 1 TO N_NEW;
    IF DIST(I,1) < DIST(I,2) THEN CL5(I) = 1;
    IF DIST(I,1) >= DIST(I,2) THEN CL5(I) = 2;
    IF CL5(I) = ACT_CL(I) THEN ERR5(I) = 0;
    ELSE ERR5(I) = 1;
END;
* PR5 = N21 | | N22 | | NT | | NIND | | H_T | | DIST | | CL5 | | ACT_CL;
* TERR5 = INT((SUM(ERR5)#/N_NEW)*100);
* PRINT PR5 TERR5;
```

*END*
*********** THIS IS FOR PROCEDURE NEIGHBOR **********;

MER1 = SUM(ER1)#/N1;
MER2 = SUM(ER2)#/N2;
MTD1 = SUM(TD1)#/N1;
MTD2 = SUM(TD2)#/N2;
*PRINT MER1 MER2 MTD1 MTD2;

S11 = (SUM((ER1'-MER1)*(ER1'-MER1'))) #(N1-1);
S21 = (SUM((TD1'-MTD1)*(TD1'-MTD1'))) #(N1-1);
COV1 = (SUM((ER1'-MER1)*(TD1'-MTD1'))) #(N1-1);
S12 = (SUM((ER2'-MER2)*(ER2'-MER2'))) #(N2-1);
S22 = (SUM((TD2'-MTD2)*(TD2'-MTD2'))) #(N2-1);
COV2 = (SUM((ER2'-MER2)*(TD2'-MTD2'))) #(N2-1);
T1 = (S11 | COV1) // (COV1 | S21);
T2 = (S12 | COV2) // (COV2 | S22);
S1P = ((N1-1)*S11 + (N2-1)*S12)#/(N1+N2-2);
S2P = ((N1-1)*S21 + (N2-1)*S22)#/(N1+N2-2);
COVP = ((N1-1)*COV1 + (N2-1)*COV2)#/(N1+N2-2);
TP = (S1P | COVP) // (COVP | S2P);
TPINV = INV(TP);
*PRINT T1 T2 TP TPINV;

DIST11 = J(N1,1,0);
DIST22 = J(N2,1,0);
NDIST11 = J(N_NEW,1,0);
NDIST22 = J(N_NEW,1,0);

**COMPUTE THE MINIMUM DISTANCE IN GROUP 1**;
DO I=1 TO N_NEW;
    DIFF11 = J(N1,1,0);
    DIFF21 = J(N1,1,0);
    DIS11 = J(N1,2,0);
    DO J=1 TO N1;
        DIFF11(J,) = ER1(J,) - NZ1(I,);
        DIFF21(J,) = TD1(J,) - NZ2(I,);
        DIS11(J,) = DIFF11(J,) | | DIFF21(J,);
        DIST11(J,) = DIS11(J,) * TPINV * DIS11(J,);
    END;
    NDIST11(I,) = MIN(DIST11);
END;

**COMPUTE THE MINIMUM DISTANCE IN GROUP 2**;
DO I=1 TO N_NEW;
    DIFF12 = J(N2,1,0);
    DIFF22 = J(N2,1,0);
    DIS22 = J(N2,2,0);
DO J=1 TO N2;
    DIFF12(J,) = ER2(J,) - NZ1(I,);
    DIFF22(J,) = TD2(J,) - NZ2(I,);
    DIS22(J,) = DIFF12(J,) || DIFF22(J,);
    DIST22(J,) = DIS22(J,) * TPINV * DIS22(J,);
END;
NDIST22(I,) = MIN(DIST22);
END;

DISTN = NDIST11 || NDIST22;

***COMPUTE THE MINIMUM OF THE DISTANCES IN GR 1 & GR 2***:
DO I = 1 TO N_NEW;
    IF DISTN(I,1) < DISTN(I,2) THEN CLN(I,) = 1;
    IF DISTN(I,1) >= DISTN(I,2) THEN CLN(I,) = 2;
    IF CLN(I,) = ACT_CL(I,) THEN ERRN(I,) = 0;
    ELSE ERRN(I,) = 1;
END;
* FIN = NZ1 || NZ2 || DISTN || CLN || ACT_CL;
TERRN = INT((SUM(ERRN)/#N_NEW/#100));
* PRINT FIN TERRN;

*****************************************************************************:

********************** CONCATENATE DIFFERENT ERROR RATES**********************:
ERROR = TERR1 || TERR2 || TERR3 || TERR4 || TERR5 || TERRN;
ERROR_2 = ERROR ## 2;
TERR(SIM,) = TERR(SIM,) + ERROR;
TERR_2(SIM,) = TERR_2(SIM,) + ERROR_2;
* PRINT ERROR ERROR_2 TERR TERR_2;
*****************************************************************************:

OUT: END;
*****END OF SIMULATION*****:

NNOSIM = NOSIM - FLAG;
NCENS = SUM(CENS) #/ NNOSIM;
PRINT FLAG NCENS;

MNPAR1 = SUM(FNPAR(,1)) #/ NNOSIM;
MNPAR2 = SUM(FNPAR(,2)) #/ NNOSIM;
MNPAR3 = SUM(FNPAR(,3)) #/ NNOSIM;
MNPAR4 = SUM(FNPAR(,4)) #/ NNOSIM;
MNPAR5 = SUM(FNPAR(,5)) #/ NNOSIM;
MNPAR6 = SUM(FNPAR(,6)) #/ NNOSIM;
FINPAR = MNPAR1 / MNPAR2 / MNPAR3 / MNPAR4 / MNPAR5 / MNPAR6;
PRINT FINPAR;
MVAR1 = SUM(FVAR(1)) #/ NNOSIM;
MVAR2 = SUM(FVAR(2)) #/ NNOSIM;
MVAR3 = SUM(FVAR(3)) #/ NNOSIM;
MVAR4 = SUM(FVAR(4)) #/ NNOSIM;
MVAR5 = SUM(FVAR(5)) #/ NNOSIM;
MVAR6 = SUM(FVAR(6)) #/ NNOSIM;
FINVAR = MVAR1 // MVAR2 // MVAR3 //
       MVAR4 // MVAR5 // MVAR6;
PRINT FINVAR;

MINERR = MIN(TERR(1)) || MIN(TERR(2)) || MIN(TERR(3)) ||
         MIN(TERR(4)) || MIN(TERR(5)) || MIN(TERR(6));

MAXERR = MAX(TERR(1)) || MAX(TERR(2)) || MAX(TERR(3)) ||
         MAX(TERR(4)) || MAX(TERR(5)) || MAX(TERR(6));

TERR1 = SUM(TERR(1));
TERR2 = SUM(TERR(2));
TERR3 = SUM(TERR(3));
TERR4 = SUM(TERR(4));
TERR5 = SUM(TERR(5));
TERRN = SUM(TERR(6));

TERR1_2 = SUM(TERR_2(1));
TERR2_2 = SUM(TERR_2(2));
TERR3_2 = SUM(TERR_2(3));
TERR4_2 = SUM(TERR_2(4));
TERR5_2 = SUM(TERR_2(5));
TERRN_2 = SUM(TERR_2(6));

MERR1 = TERR1#/NNOSIM;
MERR2 = TERR2#/NNOSIM;
MERR3 = TERR3#/NNOSIM;
MERR4 = TERR4#/NNOSIM;
MERR5 = TERR5#/NNOSIM;
MERRN = TERRN#/NNOSIM;
MEANERR = MERR1 || MERR2 || MERR3 || MERR4 || MERR5 || MERRN;

SD1=SQRT((TERR1_2-TERR1#/NNOSIM)#/(NNOSIM-1));
SD2=SQRT((TERR2_2-TERR2#/NNOSIM)#/(NNOSIM-1));
SD3=SQRT((TERR3_2-TERR3#/NNOSIM)#/(NNOSIM-1));
SD4=SQRT((TERR4_2-TERR4#/NNOSIM)#/(NNOSIM-1));
SD5=SQRT((TERR5_2-TERR5#/NNOSIM)#/(NNOSIM-1));
SDN=SQRT((TERRN_2-TERRN#/NNOSIM)#/(NNOSIM-1));
SD = SD1 || SD2 || SD3 || SD4 || SD5 || SDN;

PRINT TERR MEANERR SD MINERR MAXERR;
MR12 = SUM(COR12) #/ NNOSIM;
MR13 = SUM(COR13) #/ NNOSIM;
MR14 = SUM(COR14) #/ NNOSIM;
MR23 = SUM(COR23) #/ NNOSIM;
MR24 = SUM(COR24) #/ NNOSIM;
MR34 = SUM(COR34) #/ NNOSIM;
FINR = (1 | | MR12 | MR13 | | MR14) //
       (MR12 | 1 | | MR23 | | MR24) //
       (MR13 | MR23 | 1 | | MR34) //
       (MR14 | | MR24 | | MR34 | | 1 ) ;
PRINT FINR;

/*

*****************************************************************************************
END OF PROGRAM***********************************************************************;
## APPENDIX B

Classification Error Rates (in Percent) from Six Different Classification Procedures (Covariates: ER and Tumor Diameter; End Point: Recurrence)

<table>
<thead>
<tr>
<th>Run #</th>
<th>Classification 1</th>
<th>Classification 2</th>
<th>Classification 3</th>
<th>Classification 4</th>
<th>Classification 5</th>
<th>Neighborhood</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>40</td>
<td>40</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>40</td>
</tr>
<tr>
<td>2</td>
<td>25</td>
<td>45</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>45</td>
</tr>
<tr>
<td>3</td>
<td>50</td>
<td>35</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>45</td>
</tr>
<tr>
<td>4</td>
<td>45</td>
<td>35</td>
<td>50</td>
<td>45</td>
<td>45</td>
<td>40</td>
</tr>
<tr>
<td>5</td>
<td>50</td>
<td>55</td>
<td>35</td>
<td>30</td>
<td>30</td>
<td>50</td>
</tr>
<tr>
<td>6</td>
<td>30</td>
<td>35</td>
<td>30</td>
<td>25</td>
<td>25</td>
<td>40</td>
</tr>
<tr>
<td>7</td>
<td>30</td>
<td>45</td>
<td>20</td>
<td>15</td>
<td>15</td>
<td>30</td>
</tr>
<tr>
<td>8</td>
<td>45</td>
<td>50</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>50</td>
</tr>
<tr>
<td>9</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>40</td>
<td>30</td>
<td>35</td>
</tr>
<tr>
<td>10</td>
<td>25</td>
<td>50</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>50</td>
</tr>
<tr>
<td>11</td>
<td>40</td>
<td>55</td>
<td>30</td>
<td>30</td>
<td>35</td>
<td>55</td>
</tr>
<tr>
<td>12</td>
<td>35</td>
<td>70</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>65</td>
</tr>
<tr>
<td>13</td>
<td>40</td>
<td>45</td>
<td>30</td>
<td>25</td>
<td>25</td>
<td>50</td>
</tr>
<tr>
<td>14</td>
<td>50</td>
<td>35</td>
<td>10</td>
<td>15</td>
<td>10</td>
<td>50</td>
</tr>
<tr>
<td>15</td>
<td>40</td>
<td>25</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>25</td>
</tr>
<tr>
<td>16</td>
<td>35</td>
<td>55</td>
<td>55</td>
<td>45</td>
<td>45</td>
<td>40</td>
</tr>
<tr>
<td>17</td>
<td>60</td>
<td>45</td>
<td>35</td>
<td>40</td>
<td>40</td>
<td>50</td>
</tr>
<tr>
<td>18</td>
<td>40</td>
<td>65</td>
<td>35</td>
<td>30</td>
<td>30</td>
<td>55</td>
</tr>
<tr>
<td>19</td>
<td>40</td>
<td>65</td>
<td>45</td>
<td>45</td>
<td>40</td>
<td>60</td>
</tr>
<tr>
<td>20</td>
<td>40</td>
<td>65</td>
<td>20</td>
<td>25</td>
<td>20</td>
<td>60</td>
</tr>
<tr>
<td>21</td>
<td>60</td>
<td>55</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td>60</td>
</tr>
<tr>
<td>22</td>
<td>75</td>
<td>35</td>
<td>50</td>
<td>45</td>
<td>45</td>
<td>45</td>
</tr>
<tr>
<td>23</td>
<td>45</td>
<td>35</td>
<td>20</td>
<td>20</td>
<td>15</td>
<td>35</td>
</tr>
<tr>
<td>24</td>
<td>45</td>
<td>35</td>
<td>20</td>
<td>35</td>
<td>35</td>
<td>50</td>
</tr>
<tr>
<td>25</td>
<td>40</td>
<td>50</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>50</td>
</tr>
<tr>
<td>26</td>
<td>35</td>
<td>65</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>65</td>
</tr>
<tr>
<td>27</td>
<td>65</td>
<td>45</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>45</td>
</tr>
<tr>
<td>28</td>
<td>40</td>
<td>55</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>50</td>
</tr>
<tr>
<td>29</td>
<td>55</td>
<td>60</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>80</td>
</tr>
<tr>
<td>30</td>
<td>55</td>
<td>45</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>50</td>
</tr>
<tr>
<td>31</td>
<td>40</td>
<td>50</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>50</td>
</tr>
<tr>
<td>32</td>
<td>40</td>
<td>45</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>50</td>
</tr>
<tr>
<td>33</td>
<td>40</td>
<td>30</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>35</td>
</tr>
<tr>
<td>34</td>
<td>35</td>
<td>55</td>
<td>35</td>
<td>40</td>
<td>40</td>
<td>55</td>
</tr>
<tr>
<td>35</td>
<td>20</td>
<td>40</td>
<td>30</td>
<td>20</td>
<td>15</td>
<td>40</td>
</tr>
<tr>
<td>36</td>
<td>65</td>
<td>60</td>
<td>45</td>
<td>35</td>
<td>40</td>
<td>60</td>
</tr>
<tr>
<td>37</td>
<td>55</td>
<td>65</td>
<td>50</td>
<td>45</td>
<td>45</td>
<td>70</td>
</tr>
<tr>
<td>38</td>
<td>50</td>
<td>55</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>55</td>
</tr>
</tbody>
</table>
### Appendix B (Continued)

<table>
<thead>
<tr>
<th>Run #</th>
<th>Classification Procedure</th>
<th>Neighborhood</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>39</td>
<td>25</td>
<td>50</td>
</tr>
<tr>
<td>40</td>
<td>40</td>
<td>50</td>
</tr>
<tr>
<td>41</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>42</td>
<td>40</td>
<td>45</td>
</tr>
<tr>
<td>43</td>
<td>30</td>
<td>35</td>
</tr>
<tr>
<td>44</td>
<td>40</td>
<td>30</td>
</tr>
<tr>
<td>45</td>
<td>55</td>
<td>60</td>
</tr>
<tr>
<td>46</td>
<td>55</td>
<td>60</td>
</tr>
<tr>
<td>47</td>
<td>35</td>
<td>45</td>
</tr>
<tr>
<td>48</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>49</td>
<td>35</td>
<td>60</td>
</tr>
<tr>
<td>50</td>
<td>55</td>
<td>40</td>
</tr>
<tr>
<td>51</td>
<td>30</td>
<td>50</td>
</tr>
<tr>
<td>52</td>
<td>30</td>
<td>25</td>
</tr>
<tr>
<td>53</td>
<td>35</td>
<td>50</td>
</tr>
<tr>
<td>54</td>
<td>55</td>
<td>45</td>
</tr>
<tr>
<td>55</td>
<td>35</td>
<td>45</td>
</tr>
<tr>
<td>56</td>
<td>30</td>
<td>45</td>
</tr>
<tr>
<td>57</td>
<td>25</td>
<td>35</td>
</tr>
<tr>
<td>58</td>
<td>45</td>
<td>45</td>
</tr>
<tr>
<td>59</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>60</td>
<td>30</td>
<td>45</td>
</tr>
<tr>
<td>61</td>
<td>55</td>
<td>50</td>
</tr>
<tr>
<td>62</td>
<td>45</td>
<td>50</td>
</tr>
<tr>
<td>63</td>
<td>50</td>
<td>55</td>
</tr>
<tr>
<td>64</td>
<td>35</td>
<td>40</td>
</tr>
<tr>
<td>65</td>
<td>65</td>
<td>40</td>
</tr>
<tr>
<td>66</td>
<td>40</td>
<td>50</td>
</tr>
<tr>
<td>67</td>
<td>50</td>
<td>55</td>
</tr>
<tr>
<td>68</td>
<td>20</td>
<td>60</td>
</tr>
<tr>
<td>69</td>
<td>55</td>
<td>50</td>
</tr>
<tr>
<td>70</td>
<td>35</td>
<td>45</td>
</tr>
<tr>
<td>71</td>
<td>45</td>
<td>25</td>
</tr>
<tr>
<td>72</td>
<td>35</td>
<td>45</td>
</tr>
<tr>
<td>73</td>
<td>30</td>
<td>45</td>
</tr>
<tr>
<td>74</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>75</td>
<td>20</td>
<td>60</td>
</tr>
<tr>
<td>76</td>
<td>50</td>
<td>40</td>
</tr>
<tr>
<td>77</td>
<td>35</td>
<td>50</td>
</tr>
<tr>
<td>78</td>
<td>45</td>
<td>35</td>
</tr>
<tr>
<td>79</td>
<td>35</td>
<td>40</td>
</tr>
<tr>
<td>80</td>
<td>50</td>
<td>50</td>
</tr>
</tbody>
</table>
### Appendix B (Continued)

<table>
<thead>
<tr>
<th>Run #</th>
<th>Classification Procedure</th>
<th>Neighborhood</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>81</td>
<td>65</td>
<td>40</td>
</tr>
<tr>
<td>82</td>
<td>40</td>
<td>35</td>
</tr>
<tr>
<td>83</td>
<td>40</td>
<td>45</td>
</tr>
<tr>
<td>84</td>
<td>60</td>
<td>35</td>
</tr>
<tr>
<td>85</td>
<td>55</td>
<td>50</td>
</tr>
<tr>
<td>86</td>
<td>40</td>
<td>45</td>
</tr>
<tr>
<td>87</td>
<td>65</td>
<td>55</td>
</tr>
<tr>
<td>88</td>
<td>35</td>
<td>55</td>
</tr>
<tr>
<td>89</td>
<td>25</td>
<td>45</td>
</tr>
<tr>
<td>90</td>
<td>30</td>
<td>70</td>
</tr>
<tr>
<td>91</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>92</td>
<td>40</td>
<td>55</td>
</tr>
<tr>
<td>93</td>
<td>75</td>
<td>45</td>
</tr>
<tr>
<td>94</td>
<td>45</td>
<td>55</td>
</tr>
<tr>
<td>95</td>
<td>40</td>
<td>60</td>
</tr>
<tr>
<td>96</td>
<td>50</td>
<td>40</td>
</tr>
<tr>
<td>97</td>
<td>45</td>
<td>60</td>
</tr>
<tr>
<td>98</td>
<td>45</td>
<td>50</td>
</tr>
<tr>
<td>99</td>
<td>45</td>
<td>45</td>
</tr>
<tr>
<td>100</td>
<td>25</td>
<td>45</td>
</tr>
</tbody>
</table>
**APPENDIX C**

Classification Error Rates (in Percent) from Six Different Classification Procedures (Covariates: ER and Tumor Diameter; End Point: Breast Cancer Death)

<table>
<thead>
<tr>
<th>Run #</th>
<th>Classification Procedure 1</th>
<th>Classification Procedure 2</th>
<th>Classification Procedure 3</th>
<th>Classification Procedure 4</th>
<th>Classification Procedure 5</th>
<th>Neighborhood</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50</td>
<td>60</td>
<td>45</td>
<td>45</td>
<td>40</td>
<td>60</td>
</tr>
<tr>
<td>2</td>
<td>40</td>
<td>55</td>
<td>40</td>
<td>40</td>
<td>45</td>
<td>50</td>
</tr>
<tr>
<td>3</td>
<td>55</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>45</td>
</tr>
<tr>
<td>4</td>
<td>40</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>40</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
<td>35</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>25</td>
</tr>
<tr>
<td>6</td>
<td>60</td>
<td>35</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>40</td>
</tr>
<tr>
<td>7</td>
<td>50</td>
<td>60</td>
<td>50</td>
<td>35</td>
<td>35</td>
<td>70</td>
</tr>
<tr>
<td>8</td>
<td>15</td>
<td>35</td>
<td>20</td>
<td>35</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>9</td>
<td>70</td>
<td>50</td>
<td>55</td>
<td>55</td>
<td>60</td>
<td>50</td>
</tr>
<tr>
<td>10</td>
<td>20</td>
<td>50</td>
<td>50</td>
<td>35</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>11</td>
<td>40</td>
<td>70</td>
<td>45</td>
<td>40</td>
<td>40</td>
<td>65</td>
</tr>
<tr>
<td>12</td>
<td>35</td>
<td>55</td>
<td>20</td>
<td>20</td>
<td>10</td>
<td>50</td>
</tr>
<tr>
<td>13</td>
<td>30</td>
<td>50</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>50</td>
</tr>
<tr>
<td>14</td>
<td>25</td>
<td>50</td>
<td>25</td>
<td>20</td>
<td>20</td>
<td>60</td>
</tr>
<tr>
<td>15</td>
<td>40</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>30</td>
<td>35</td>
</tr>
<tr>
<td>16</td>
<td>35</td>
<td>40</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>55</td>
</tr>
<tr>
<td>17</td>
<td>30</td>
<td>35</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>30</td>
</tr>
<tr>
<td>18</td>
<td>30</td>
<td>60</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>55</td>
</tr>
<tr>
<td>19</td>
<td>40</td>
<td>35</td>
<td>30</td>
<td>35</td>
<td>30</td>
<td>40</td>
</tr>
<tr>
<td>20</td>
<td>40</td>
<td>40</td>
<td>35</td>
<td>40</td>
<td>40</td>
<td>50</td>
</tr>
<tr>
<td>21</td>
<td>65</td>
<td>50</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>40</td>
</tr>
<tr>
<td>22</td>
<td>30</td>
<td>45</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>40</td>
</tr>
<tr>
<td>23</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>40</td>
<td>40</td>
<td>65</td>
</tr>
<tr>
<td>24</td>
<td>45</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td>30</td>
</tr>
<tr>
<td>25</td>
<td>60</td>
<td>30</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
</tr>
<tr>
<td>26</td>
<td>35</td>
<td>55</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td>50</td>
</tr>
<tr>
<td>27</td>
<td>50</td>
<td>55</td>
<td>25</td>
<td>35</td>
<td>20</td>
<td>65</td>
</tr>
<tr>
<td>28</td>
<td>40</td>
<td>35</td>
<td>30</td>
<td>40</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>29</td>
<td>45</td>
<td>55</td>
<td>30</td>
<td>25</td>
<td>25</td>
<td>60</td>
</tr>
<tr>
<td>30</td>
<td>30</td>
<td>30</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>30</td>
</tr>
<tr>
<td>31</td>
<td>40</td>
<td>60</td>
<td>40</td>
<td>35</td>
<td>30</td>
<td>60</td>
</tr>
<tr>
<td>32</td>
<td>40</td>
<td>45</td>
<td>35</td>
<td>30</td>
<td>35</td>
<td>45</td>
</tr>
<tr>
<td>33</td>
<td>30</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>34</td>
<td>45</td>
<td>40</td>
<td>40</td>
<td>35</td>
<td>35</td>
<td>45</td>
</tr>
<tr>
<td>35</td>
<td>45</td>
<td>55</td>
<td>35</td>
<td>40</td>
<td>40</td>
<td>55</td>
</tr>
<tr>
<td>36</td>
<td>45</td>
<td>35</td>
<td>45</td>
<td>45</td>
<td>40</td>
<td>30</td>
</tr>
<tr>
<td>37</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td>30</td>
<td>30</td>
<td>35</td>
</tr>
<tr>
<td>38</td>
<td>50</td>
<td>35</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>35</td>
</tr>
</tbody>
</table>
APPENDIX C  (Continued)

<table>
<thead>
<tr>
<th>Run #</th>
<th>Classification Procedure 1</th>
<th>Classification Procedure 2</th>
<th>Classification Procedure 3</th>
<th>Classification Procedure 4</th>
<th>Classification Procedure 5</th>
<th>Neighborhood</th>
</tr>
</thead>
<tbody>
<tr>
<td>39</td>
<td>50</td>
<td>75</td>
<td>55</td>
<td>50</td>
<td>50</td>
<td>70</td>
</tr>
<tr>
<td>40</td>
<td>45</td>
<td>65</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>60</td>
</tr>
<tr>
<td>41</td>
<td>30</td>
<td>35</td>
<td>25</td>
<td>25</td>
<td>30</td>
<td>25</td>
</tr>
<tr>
<td>42</td>
<td>35</td>
<td>30</td>
<td>35</td>
<td>35</td>
<td>30</td>
<td>35</td>
</tr>
<tr>
<td>43</td>
<td>55</td>
<td>35</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td>45</td>
</tr>
<tr>
<td>44</td>
<td>55</td>
<td>75</td>
<td>30</td>
<td>10</td>
<td>20</td>
<td>65</td>
</tr>
<tr>
<td>45</td>
<td>35</td>
<td>45</td>
<td>25</td>
<td>25</td>
<td>30</td>
<td>50</td>
</tr>
<tr>
<td>46</td>
<td>45</td>
<td>45</td>
<td>30</td>
<td>25</td>
<td>25</td>
<td>40</td>
</tr>
<tr>
<td>47</td>
<td>55</td>
<td>50</td>
<td>45</td>
<td>40</td>
<td>35</td>
<td>55</td>
</tr>
<tr>
<td>48</td>
<td>35</td>
<td>60</td>
<td>55</td>
<td>40</td>
<td>40</td>
<td>55</td>
</tr>
<tr>
<td>49</td>
<td>35</td>
<td>40</td>
<td>20</td>
<td>15</td>
<td>10</td>
<td>30</td>
</tr>
<tr>
<td>50</td>
<td>40</td>
<td>45</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>55</td>
</tr>
<tr>
<td>51</td>
<td>40</td>
<td>55</td>
<td>50</td>
<td>45</td>
<td>45</td>
<td>65</td>
</tr>
<tr>
<td>52</td>
<td>25</td>
<td>40</td>
<td>25</td>
<td>20</td>
<td>25</td>
<td>40</td>
</tr>
<tr>
<td>53</td>
<td>60</td>
<td>55</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td>60</td>
</tr>
<tr>
<td>54</td>
<td>55</td>
<td>60</td>
<td>40</td>
<td>40</td>
<td>35</td>
<td>60</td>
</tr>
<tr>
<td>55</td>
<td>45</td>
<td>50</td>
<td>50</td>
<td>45</td>
<td>45</td>
<td>60</td>
</tr>
<tr>
<td>56</td>
<td>20</td>
<td>30</td>
<td>30</td>
<td>25</td>
<td>25</td>
<td>35</td>
</tr>
<tr>
<td>57</td>
<td>50</td>
<td>60</td>
<td>20</td>
<td>15</td>
<td>15</td>
<td>55</td>
</tr>
<tr>
<td>58</td>
<td>40</td>
<td>40</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>45</td>
</tr>
<tr>
<td>59</td>
<td>45</td>
<td>35</td>
<td>30</td>
<td>25</td>
<td>25</td>
<td>45</td>
</tr>
<tr>
<td>60</td>
<td>35</td>
<td>60</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>60</td>
</tr>
<tr>
<td>61</td>
<td>55</td>
<td>30</td>
<td>30</td>
<td>35</td>
<td>35</td>
<td>45</td>
</tr>
<tr>
<td>62</td>
<td>40</td>
<td>55</td>
<td>60</td>
<td>45</td>
<td>45</td>
<td>70</td>
</tr>
<tr>
<td>63</td>
<td>45</td>
<td>75</td>
<td>35</td>
<td>30</td>
<td>30</td>
<td>70</td>
</tr>
<tr>
<td>64</td>
<td>50</td>
<td>50</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>45</td>
</tr>
<tr>
<td>65</td>
<td>45</td>
<td>75</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>65</td>
</tr>
<tr>
<td>66</td>
<td>30</td>
<td>40</td>
<td>35</td>
<td>45</td>
<td>40</td>
<td>35</td>
</tr>
<tr>
<td>67</td>
<td>40</td>
<td>55</td>
<td>60</td>
<td>45</td>
<td>45</td>
<td>60</td>
</tr>
<tr>
<td>68</td>
<td>45</td>
<td>30</td>
<td>25</td>
<td>30</td>
<td>30</td>
<td>50</td>
</tr>
<tr>
<td>69</td>
<td>45</td>
<td>35</td>
<td>30</td>
<td>35</td>
<td>40</td>
<td>45</td>
</tr>
<tr>
<td>70</td>
<td>40</td>
<td>60</td>
<td>45</td>
<td>40</td>
<td>40</td>
<td>55</td>
</tr>
<tr>
<td>71</td>
<td>40</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>40</td>
</tr>
<tr>
<td>72</td>
<td>45</td>
<td>40</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>35</td>
</tr>
<tr>
<td>73</td>
<td>20</td>
<td>55</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>55</td>
</tr>
<tr>
<td>74</td>
<td>35</td>
<td>50</td>
<td>30</td>
<td>35</td>
<td>35</td>
<td>45</td>
</tr>
<tr>
<td>75</td>
<td>40</td>
<td>50</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>55</td>
</tr>
<tr>
<td>76</td>
<td>30</td>
<td>40</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>35</td>
</tr>
<tr>
<td>77</td>
<td>40</td>
<td>50</td>
<td>25</td>
<td>30</td>
<td>30</td>
<td>45</td>
</tr>
<tr>
<td>78</td>
<td>40</td>
<td>55</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td>55</td>
</tr>
<tr>
<td>79</td>
<td>10</td>
<td>30</td>
<td>15</td>
<td>20</td>
<td>20</td>
<td>40</td>
</tr>
<tr>
<td>80</td>
<td>50</td>
<td>70</td>
<td>70</td>
<td>60</td>
<td>60</td>
<td>65</td>
</tr>
</tbody>
</table>
## APPENDIX C (Continued)

<table>
<thead>
<tr>
<th>Run #</th>
<th>Classification Procedure</th>
<th>Neighborhood</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>81</td>
<td></td>
<td>40</td>
</tr>
<tr>
<td>82</td>
<td></td>
<td>50</td>
</tr>
<tr>
<td>83</td>
<td></td>
<td>45</td>
</tr>
<tr>
<td>84</td>
<td></td>
<td>40</td>
</tr>
<tr>
<td>85</td>
<td></td>
<td>40</td>
</tr>
<tr>
<td>86</td>
<td></td>
<td>45</td>
</tr>
<tr>
<td>87</td>
<td></td>
<td>50</td>
</tr>
<tr>
<td>88</td>
<td></td>
<td>50</td>
</tr>
<tr>
<td>89</td>
<td></td>
<td>35</td>
</tr>
<tr>
<td>90</td>
<td></td>
<td>50</td>
</tr>
<tr>
<td>91</td>
<td></td>
<td>60</td>
</tr>
<tr>
<td>92</td>
<td></td>
<td>40</td>
</tr>
<tr>
<td>93</td>
<td></td>
<td>40</td>
</tr>
<tr>
<td>94</td>
<td></td>
<td>40</td>
</tr>
<tr>
<td>95</td>
<td></td>
<td>60</td>
</tr>
<tr>
<td>96</td>
<td></td>
<td>40</td>
</tr>
<tr>
<td>97</td>
<td></td>
<td>45</td>
</tr>
<tr>
<td>98</td>
<td></td>
<td>35</td>
</tr>
<tr>
<td>99</td>
<td></td>
<td>55</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td>25</td>
</tr>
</tbody>
</table>
APPENDIX D

SIMULATION PROGRAM

******************************************************************************;

PROC MATRIX;

NTOT=350;
NOTEST=50;
PP=0.50; *PP IS THE PROPORTION OF GROUP 1;
N1TOT=INT(PP*NTOT);
N2TOT=NTOT-N1TOT;
N1TEST=INT(PP*NOTEST);
N2TEST=NOTEST-N1TEST;

******************************************************************************;

***THESE ARE THE PARAMETERS TO BE VARIED***:
TL = 20;
A01 = -2.99; A11 = -0.12; A21 = 0.03;
A02 = -2.00; A12 = -0.25; A22 = 0.08;
SUCC1 = 0.99; SUCC2 = 0.05;
MU10 = 4.0; SIGM10 = 1;
MU11 = 1.0; SIGM11 = 1;
MU20 = 8.0; SIGM20 = 1;
MU21 = 2.0; SIGM21 = 1;

******************************************************************************;

NOSIM = 100;
CENS = J(NOSIM,1,0);
TERR = J(NOSIM,6,0);
TERR_2 = J(NOSIM,6,0);

COR12 = J(NOSIM,1,0); COR13 = J(NOSIM,1,0);
COR14 = J(NOSIM,1,0); COR23 = J(NOSIM,1,0);
COR24 = J(NOSIM,1,0); COR34 = J(NOSIM,1,0);
FINR = J(4,4,0);

FNPAR1 = J(NOSIM,1,0); FNPAR2 = J(NOSIM,1,0);
FNPAR3 = J(NOSIM,1,0); FNPAR4 = J(NOSIM,1,0);
FNPAR5 = J(NOSIM,1,0); FNPAR6 = J(NOSIM,1,0);

FVAR1 = J(NOSIM,1,0); FVAR2 = J(NOSIM,1,0); FVAR3 = J(NOSIM,1,0);
FVAR4 = J(NOSIM,1,0); FVAR5 = J(NOSIM,1,0); FVAR6 = J(NOSIM,1,0);
FLAG = 0;

DO SIM = 1 TO NOSIM;

*** THIS IS TO GENERATE ACCRUAL TIMES OVER U(0.5) ***;
   GA1 = 5#UNIFORM(J(N1TOT,1,0));
   GA2 = 5#UNIFORM(J(N2TOT,1,0));

*******************************************************************************;

GPN1=J(N1TOT,1,1);
GPN2=J(N2TOT,1,2);
GER1=J(N1TOT,1,0);
GER2=J(N2TOT,1,0);
GTD1=J(N1TOT,1,0);
GTD2=J(N2TOT,1,0);
GS1=J(N1TOT,1,0);
GS2=J(N2TOT,1,0);
G1D1=J(N1TOT,1,0);
G1D2=J(N2TOT,1,0);
GLL1=J(N1TOT,1,0);
GLL2=J(N2TOT,1,0);

*******************************************************************************

*** THIS IS TO GENERATE COVARIATES FOR PN = 1*******;
DO M=1 TO N1TOT;
   GER1(M,) = RANBIN(0,1,SUCC1);
   IF GER1(M,) = 0 THEN
      GTD1(M,) = ABS(INT(MU10+SIGM10*RANNOR(0)));
   IF GER1(M,) = 1 THEN
      GTD1(M,) = ABS(INT(MU11+SIGM11*RANNOR(0)));
   GLL1(M,) = EXP(A01+(A11#GER1(M,))+(A21#GTD1(M,)));
   GS1(M,) = (- LOG(UNIFORM(0))#GLL1(M,);  
END;

*******************************************************************************

*** THIS IS TO GENERATE COVARIATES FOR PN = 2*******;
DO M=1 TO N2TOT;
   GER2(M,) = RANBIN(0,1,SUCC2);
   IF GER2(M,) = 0 THEN
      GTD2(M,) = ABS(INT(MU20+SIGM20*RANNOR(0)));
   IF GER2(M,) = 1 THEN
      GTD2(M,) = ABS(INT(MU21+SIGM21*RANNOR(0)));
   GLL2(M,) = EXP(A02+(A12#GER2(M,))+(A22#GTD2(M,)));
   GS2(M,) = (- LOG(UNIFORM(0))#GLL2(M,);  
END;

*******************************************************************************
AS1 = GA1 + GS1;
AS2 = GA2 + GS2;

GT1 = J(T1TOT, 1, 0);
GIND1 = J(T1TOT, 1, 0);
GC11 = J(T1TOT, 1, 0);
GC21 = J(T1TOT, 1, 0);
GC31 = J(T1TOT, 1, 0);
GC41 = J(T1TOT, 1, 0);
GT2 = J(T2TOT, 1, 0);
GIND2 = J(T2TOT, 1, 0);
GC12 = J(T2TOT, 1, 0);
GC22 = J(T2TOT, 1, 0);
GC32 = J(T2TOT, 1, 0);
GC42 = J(T2TOT, 1, 0);

DO I = 1 TO T1TOT;
IF AS1(I) > TL THEN DO:
   GIND1(I) = 0;
   GT1(I) = TL - GA1(I);
   IF GPN1(I) = 1 THEN GC31(I) = 1;
   ELSE GC31(I) = 0;
   GC41(I) = 0;
END;
ELSE DO:
   GIND1(I) = 1;
   GT1(I) = GS1(I);
   IF GPN1(I) = 1 THEN GC11(I) = 1;
   ELSE GC11(I) = 0;
   GC21(I) = 0;
END;
END;

DO I = 1 TO T2TOT;
IF AS2(I) > TL THEN DO:
   GIND2(I) = 0;
   GT2(I) = TL - GA2(I);
   GC32(I) = 0;
   IF GPN2(I) = 2 THEN GC42(I) = 1;
   ELSE GC42(I) = 0;
END;
ELSE DO:
   GIND2(I) = 1;
   GT2(I) = GS2(I);
   GC12(I) = 0;
   IF GPN2(I) = 2 THEN GC22(I) = 1;
   ELSE GC22(I) = 0;
END;
END;
INP1 = GPN1 | GER1 | GTD1 | GT1 | GIND1 | GC11 | GC21 | GC31 | GC41;
INP2 = GPN2 | GER2 | GTD2 | GT2 | GIND2 | GC12 | GC22 | GC32 | GC42;
INP = INP1 // INP2;

***THIS IS TO EXCLUDE 50 OBSERVATIONS FROM THE ANALYSIS RANDOMLY***;
    U = INT (UNIFORM(J(NOTES,T,1,0))#NTOT)+1;
    ALL = (1:NTOT);
    ALL(U) = 0;
    NINP = INP(LOC(ALL > 0),);
    NEWINP = INP(U,);
    NINP1 = NINP(LOC(NINP(1)=1),);
    NINP2 = NINP(LOC(NINP(1)=2),);
    NINP = NINP1 // NINP2;

*******************************************************************************************;

*******************************************************************************************;

***THIS IS TO COMPUTE THE CORRELATION MATRIX***;
    X = NINP(:,1) | | NINP(:,2) | | NINP(:,3) | | NINP(:,5);
    NUMER = NROW(X);
    X_SUM = X(+,);
    XPX = X'X - X_SUM'X_SUM#/NUMER;
    SCOR = 1#/SQR(DIAG(XPX));
    CORR = SCOR*XPX*SCOR;
*
    PRINT CORR;
    COR12(SIM.) = CORR(1,2);
    COR13(SIM.) = CORR(1,3);
    COR14(SIM.) = CORR(1,4);
    COR23(SIM.) = CORR(2,3);
    COR24(SIM.) = CORR(2,4);
    COR34(SIM.) = CORR(3,4);

*******************************************************************************************;

N1 = NROW(NINP1);
N2 = NROW(NINP2);
N = N1 + N2;
RK1 = J(N1,1,0);
RK2 = J(N2,1,0);
S1_EST = J(N1,1,0);
S2_EST = J(N2,1,0);
RK1 = RANK(NINP1(,4));
RK2 = RANK(NINP2(,4));
IND1 = NINP1(,5);
IND2 = NINP2(,5);
T1 = NINP1(,4);
T2 = NINP2(,4);
PN1 = NINP1(,1);
PN2 = NINP2(,1);
ER1 = NINP1(,2);
ER2 = NINP2(,2);
TD1 = NINP1(3);
TD2 = NINP2(3);
L1 = J(N1,1,0);
L2 = J(N2,1,0);
X01 = J(N1,1,1);
X02 = J(N2,1,1);
PAR1 = J(3,1,0);
PAR2 = J(3,1,0);

DO M=1 TO N1;
    IF IND1(M,) = 1 THEN S1_EST(M,) = (N1 - RK1(M,) + 1)#/(N1+1);
    ELSE S1_EST(M,) = 0;
    IF S1_EST(M,) = 0 THEN L1(M,) = LOG(( - LOG(S1_EST(M,)))/T1(M,));
    ELSE L1(M,) = 0;
    IF L1(M,) = 0 THEN DO;
        ER1(M,) = 0;
        TD1(M,) = 0;
        X01(M,) = 0;
    END;
END;

DO M=1 TO N2;
    IF IND2(M,) = 1 THEN S2_EST(M,) = (N2 - RK2(M,) + 1)#/(N2+1);
    ELSE S2_EST(M,) = 0;
    IF S2_EST(M,) = 0 THEN L2(M,) = LOG(( - LOG(S2_EST(M,)))/T2(M,));
    ELSE L2(M,) = 0;
    IF L2(M,) = 0 THEN DO;
        ER2(M,) = 0;
        TD2(M,) = 0;
        X02(M,) = 0;
    END;
END;

TEMP1 = L1 || X01 | | ER1 | | TD1;
TEMP2 = L2 || X02 | | ER2 | | TD2;
NTEMP1 = TEMP1(LOC(TEMP1(1),)=0,);
NTEMP2 = TEMP2(LOC(TEMP2(1),)=0,);
X1 = NTEMP1(2) || NTEMP1(3) || NTEMP1(4);
X2 = NTEMP2(2) || NTEMP2(3) || NTEMP2(4);
XX1 = X1' * X1;
XX2 = X2' * X2;
IF DET(XX1) = 0 OR DET(XX2) = 0 THEN DO;
    FLAG = FLAG + 1;
    GOTO OUT;
END;
ELSE GOTO NOW;
NOW: INVX1 = INV(XX1);
INVX2 = INV(XX2);
PAR1 = INVX1 * X1' * NTEMP1(1); PAR2 = INVX2 * X2' * NTEMP2(1);
PAR = PAR1 // PAR2;
*****THIS IS THE NEWTON-RAPHSON PROCEDURE*****;

NPAR = PAR;
Z1 = NINP(.2); Z2 = NINP(.3); T = NINP(.4);
C1 = NINP(.6); C2 = NINP(.7);
C3 = NINP(.8); C4 = NINP(.9);
CC1 = SUM(C1); CC2 = SUM(C2);
CC3 = SUM(C3); CC4 = SUM(C4);
N = CC1 + CC2 + CC3 + CC4;
P1 = CC1/#N; P2 = CC2/#N;
P3 = CC3/#N; P4 = 1 - P1 - P2 - P3;
U = J(6,1,0);
I = J(6,6,0);
COUNT = 0;
BEGIN: PAR = NPAR;
    B01 = PAR(1.);
    B11 = PAR(2.);
    B21 = PAR(3.);
    B02 = PAR(4.);
    B12 = PAR(5.);
    B22 = PAR(6.);

LL = SUM(C1#(LOG(P1)+B01+B11#Z1+B21#Z2-
    T#(EXP(B01+B11#Z1+B21#Z2)))+
    C2#(LOG(P2)+B02+B12#Z1+B22#Z2-
    T#(EXP(B02+B12#Z1+B22#Z2)))+
    C3#(LOG(P3)-T#(EXP(B01+B11#Z1+B21#Z2)))+
    C4#(LOG(P4)-T#(EXP(B02+B12#Z1+B22#Z2))));

**THESE ARE THE FIRST DERIVATIVES OF THE LOG-LIKELIHOOD
FUNCTION**;

U(1.) = SUM((C1#((1-(EXP(B01+B11#Z1+B21#Z2))/#T)-
    C3##((EXP(B01+B11#Z1+B21#Z2))/#T)));

U(2.) = SUM((C1#(Z1-Z1#T#(EXP(B01+B11#Z1+B21#Z2)))-
    C3#(Z1#T#(EXP(B01+B11#Z1+B21#Z2))));

U(3.) = SUM((C1#(Z2-Z2#T#(EXP(B01+B11#Z1+B21#Z2)))-
    C3#(Z2#T#(EXP(B01+B11#Z1+B21#Z2))));

U(4.) = SUM((C2#((1-(EXP(B02+B12#Z1+B22#Z2))/#T)-
    C4##((EXP(B02+B12#Z1+B22#Z2))/#T)));

U(5.) = SUM((C2#(Z1-Z1#T#(EXP(B02+B12#Z1+B22#Z2)))-
    C4#(Z1#T#(EXP(B02+B12#Z1+B22#Z2))));

U(6.) = SUM((C2#(Z2-Z2#T#(EXP(B02+B12#Z1+B22#Z2)))-
    C4#(Z2#T#(EXP(B02+B12#Z1+B22#Z2))));
**THESE ARE THE SECOND DERIVATIVES OF THE LOG-LIKELIHOOD***

**FUNCTION**

\[ I(1,1) = \text{SUM}(-C1#((\text{EXP}(B01+B11#Z1+B21#Z2))#T)-C3#((\text{EXP}(B01+B11#Z1+B21#Z2))#T)); \]

\[ I(1,2) = \text{SUM}(-C1#(Z1#T#(\text{EXP}(B01+B11#Z1+B21#Z2)))-C3#(Z1#T#(\text{EXP}(B01+B11#Z1+B21#Z2)))); \]

\[ I(1,3) = \text{SUM}(-C1#(Z2#T#(\text{EXP}(B01+B11#Z1+B21#Z2)))-C3#(Z2#T#(\text{EXP}(B01+B11#Z1+B21#Z2)))); \]

\[ I(1,4) = 0; \quad I(1,5) = 0; \quad I(1,6) = 0; \]

\[ I(2,1) = I(1,2); \]

\[ I(2,2) = \text{SUM}(-C1#(Z1#Z1#T#(\text{EXP}(B01+B11#Z1+B21#Z2)))-C3#(Z1#Z1#T#(\text{EXP}(B01+B11#Z1+B21#Z2)))); \]

\[ I(2,3) = \text{SUM}(-C1#(Z1#Z2#T#(\text{EXP}(B01+B11#Z1+B21#Z2)))-C3#(Z1#Z2#T#(\text{EXP}(B01+B11#Z1+B21#Z2)))); \]

\[ I(2,4) = 0; \quad I(2,5) = 0; \quad I(2,6) = 0; \]

\[ I(3,1) = I(1,3); \quad I(3,2) = I(2,3); \]

\[ I(3,3) = \text{SUM}(-C1#(Z2#Z2#T#(\text{EXP}(B01+B11#Z1+B21#Z2)))-C3#(Z2#Z2#T#(\text{EXP}(B01+B11#Z1+B21#Z2)))); \]

\[ I(3,4) = 0; \quad I(3,5) = 0; \quad I(3,6) = 0; \]

\[ I(4,1) = I(1,4); \quad I(4,2) = I(2,4); \quad I(4,3) = I(3,4); \]

\[ I(4,4) = \text{SUM}(-C2#((\text{EXP}(B02+B12#Z1+B22#Z2))#T)-C4#((\text{EXP}(B02+B12#Z1+B22#Z2))#T)); \]

\[ I(4,5) = \text{SUM}(-C2#(Z1#T#(\text{EXP}(B02+B12#Z1+B22#Z2)))-C4#(Z1#T#(\text{EXP}(B02+B12#Z1+B22#Z2)))); \]

\[ I(4,6) = \text{SUM}(-C2#(Z2#T#(\text{EXP}(B02+B12#Z1+B22#Z2)))-C4#(Z2#T#(\text{EXP}(B02+B12#Z1+B22#Z2)))); \]

\[ I(5,1) = I(1,5); \quad I(5,2) = I(2,5); \quad I(5,3) = I(3,5); \quad I(5,4) = I(4,5); \]

\[ I(5,5) = \text{SUM}(-C2#(Z1#Z1#T#(\text{EXP}(B02+B12#Z1+B22#Z2)))-C4#(Z1#Z1#T#(\text{EXP}(B02+B12#Z1+B22#Z2)))); \]

\[ I(5,6) = \text{SUM}(-C2#(Z1#Z2#T#(\text{EXP}(B02+B12#Z1+B22#Z2)))-C4#(Z1#Z2#T#(\text{EXP}(B02+B12#Z1+B22#Z2)))); \]

\[ I(6,1) = I(1,6); \quad I(6,2) = I(2,6); \quad I(6,3) = I(3,6); \quad I(6,4) = I(4,6); \quad I(6,5) = I(5,6); \]
\[
I(6,6) = \text{SUM}(-C2\#(Z2\#Z2\#T\#(\text{EXP}(B02+B12\#Z1+B22\#Z2)) - \\
C4\#(Z2\#Z2\#T\#(\text{EXP}(B02+B12\#Z1+B22\#Z2)))));
\]

IF DET(I) = 0 THEN DO;
   FLAG = FLAG+1;
   GOTO OUT;
END;
ELSE GOTO CONT;
CONT: INV_ = INV(I);
NPAR = PAR_ - INV_*U;
* PRINT LL U I INV_ NPAR;
COUNT = COUNT + 1;
IF COUNT < 10 THEN GOTO BEGIN;
ELSE GOTO FIN;
FIN: D = DET(INV_);
PRINT LL D I INV_ NPAR;

FNPAR1(SIM_1) = NPAR(1,);
FNPAR2(SIM_1) = NPAR(2,);
FNPAR3(SIM_1) = NPAR(3,);
FNPAR4(SIM_1) = NPAR(4,);
FNPAR5(SIM_1) = NPAR(5,);
FNPAR6(SIM_1) = NPAR(6,);

FVAR1(SIM_1) = INV(1,1);
FVAR2(SIM_1) = INV(2,2);
FVAR3 = INV(3,3);
FVAR4(SIM_1) = INV(4,4);
FVAR5(SIM_1) = INV(5,5);
FVAR6 = INV(6,6);

**** RECALCULATE HAZARD FOR EACH GROUP ****

\[
\begin{align*}
HZ1 &= (C1 \# (\text{EXP}(NPAR(1,)+NPAR(2,)\#Z1+NPAR(3,)\#Z2))) + \\
     & \quad (C3 \# (\text{EXP}(NPAR(1,)+NPAR(2,)\#Z1+NPAR(3,)\#Z2)) ; \\
HZ2 &= (C2 \# (\text{EXP}(NPAR(4,)+NPAR(5,)\#Z1+NPAR(6,)\#Z2))) + \\
     & \quad (C4 \# (\text{EXP}(NPAR(4,)+NPAR(5,)\#Z1+NPAR(6,)\#Z2)) ; \\
\end{align*}
\]

H1 = HZ1(LOC(HZ1>0,));
H2 = HZ2(LOC(HZ2>0,));
HH1 = H1; HH2 = H2;
H1(RANK(H1,)) = HH1;
H2(RANK(H2,)) = HH2;

MH1 = SUM(H1) \#/ N1;
MH2 = SUM(H2) \#/ N2;
S11 = (SUM((H1'-MH1)\*(H1'-MH1)')) \#/ (N1 - 1);
S22 = (SUM((H2'-MH2)\*(H2'-MH2)')) \#/ (N2 - 1);
SP = ((N1-1)\*S11 + (N2-1)\*S22) \#/ (N1+N2-2);
STDP = SQRT(SP);
*PRINT MH1 MH2 S11 S22 SP STDP;

PC1 = P1 + P3; *THIS IS THE PRIOR PROBABILITY OF BEING IN CLASS 1:
PC2 = P2 + P4; *THIS IS THE PRIOR PROBABILITY OF BEING IN CLASS 2:
CENS(SIM,) = P3 + P4; * THIS IS THE PROPORTION CENSORED; *
PRINT P1 P2 P3 P4 PC1 PC2 CENS;

NIND1 = J(NOTE,1,0);
NIND2 = J(NOTE,1,0);
NIND3 = J(NOTE,1,0);
NIND4 = J(NOTE,1,0);
ACT_CL = NEWINP(1);
NZ1 = NEWINP(2);
NZ2 = NEWINP(3);
NT = NEWINP(4);
NIND = NEWINP(5);
DO W = 1 TO NOTE;
  IF NIND(W) = 1 THEN DO;
    NIND1(W) = 1; NIND2(W) = 1;
    NIND3(W) = 0; NIND4(W) = 0;
  END;
  IF NIND(W) = 0 THEN DO;
    NIND1(W) = 0; NIND2(W) = 0;
    NIND3(W) = 1; NIND4(W) = 1;
  END;
END;

H1NEW = EXP(NPAR(1,1)+(NPAR(2,1)#NZ1)+(NPAR(3,1)#NZ2));
H2NEW = EXP(NPAR(4,1)+(NPAR(5,1)#NZ1)+(NPAR(6,1)#NZ2));

F1 = H1NEW # (EXP(-H1NEW#NT));
F2 = H2NEW # (EXP(-H2NEW#NT));

S1 = EXP(-H1NEW#NT);
S2 = EXP(-H2NEW#NT);
* FMIX = (P1#F1)+(P2#F2);
* SMIX = (P3#S1)+(P4#S2);

P1_T = ((NIND1#(P1#F1))+(NIND3#(P3#S1))) #/
       ((NIND1#(P1#F1)+P2#F2))+(NIND3#(P3#S1)+P4#S2));
P2_T = ((NIND2#(P2#F2))+(NIND4#(P4#S2))) #/
       ((NIND2#(P1#F1)+P2#F2))+(NIND4#(P3#S1)+P4#S2));
F_T = P1_T#F1 + P2_T#F2;
S_T = P1_T#S1 + P2_T#S2;
H_T = F_T #/ S_T;

CL1 = J(NOTE,1,0);
CL2 = J(NOTE,1,0);
CL3 = J(NOTE,1,0);
CL4 = J(NOTE,1,0);
CL5 = J(NOTE,1,0);
CLN = J(NOTE,1,0);
EST PRI = J(NOTE,1,0);
ESTPR2 = J(NOTEST,1,0);
PR1A = J(NOTEST,1,0);
PR2A = J(NOTEST,1,0);
PR3A = J(NOTEST,1,0);
PR4A = J(NOTEST,1,0);
ERR1 = J(NOTEST,1,0);
ERR2 = J(NOTEST,1,0);
ERR3 = J(NOTEST,1,0);
ERR4 = J(NOTEST,1,0);
ERR5 = J(NOTEST,1,0);
ERRN = J(NOTEST,1,0);
DIFF1 = J(N1,1,0);
DIFF2 = J(N2,1,0);
DIST1 = J(N1,1,0);
DIST2 = J(N2,1,0);
NDIST1 = J(NOTEST,1,0);
NDIST2 = J(NOTEST,1,0);

******************************************************************************;
*THIS IS FOR CLASSIFICATION METHOD 1 (BASED ON 2 HAZARD VALUES);
******************************************************************************;

DO J = 1 TO NOTEST;
  K = 1;
  DO K = 1 TO N1;
    IF H1NEW(J) <= H1(K) THEN GOTO COUNT1;
  END;
  COUNT1: PR1A(J) = (N1-(K-1))#/N1;
END;

DO J = 1 TO NOTEST;
  K = 1;
  DO K = 1 TO N2;
    IF H2NEW(J) < H2(K) THEN GOTO COUNT4;
  END;
  COUNT4: PR4A(J) = (K-1)#/N2;
END;

DO L=1 TO NOTEST;
  IF PR1A(L) > PR4A(L) THEN CL1(L) = 1;
  IF PR1A(L) <= PR4A(L) THEN CL1(L) = 2;
  IF CL1(L)=ACT_CL(L) THEN ERR1(L)=0;
  ELSE ERR1(L) = 1;
END;

* PR1 = NZ1 || NZ2 || NT || NIND || PR1A || PR4A || CL1 || ACT_CL;
TERR1 = INT((SUM(ERR1)#/NOTEST)*100);
* PRINT PR1 TERR1;

******************************************************************************:
DO J = 1 TO NOTEST;
    K = 1;
    DO K = 1 TO N1;
        IF H2NEW(J,) <= H1(K,) THEN GOTO COUNT2;
        END;
        COUNT2: PR2A(J,) = (N1-(K-1))#/N1;
    END;
    DO J = 1 TO NOTEST;
    K = 1;
    DO K = 1 TO N2;
        IF H1NEW(J,) < H2(K,) THEN GOTO COUNT3;
        END;
        COUNT3: PR3A(J,) = (K-1)#/N2;
    END;
    PROM1= J(NOTEST,1,0);
    PROM2= J(NOTEST,1,0);
    PRO1 = J(NOTEST,1,0);
    PRO2 = J(NOTEST,1,0);
    DO L=1 TO NOTEST;
        PROM1(L,) = (PR1A(L,)##PC1) + (PR3A(L,)##PC2);
        PROM2(L,) = (PR2A(L,)##PC1) + (PR4A(L,)##PC2);
        PRO1(L,) = (PR1A(L,)##PC1) #/ PROM1(L,);
        PRO2(L,) = (PR4A(L,)##PC2) #/ PROM2(L,);
        IF PRO1(L,) > PRO2(L,) THEN CL2(L,) = 1;
        IF PRO1(L,) <= PRO2(L,) THEN CL2(L,) = 2;
        IF CL2(L,) = ACT.CL(L,) THEN ERR2(L,) = 0;
        ELSE ERR2(L,) = 1;
    END;
    * PR2 = NZ1 | | NZ2 | | NT | | NIND | | PRO1 | | PRO2 | | CL2 | | ACT.CL:
        TERR2 = INT((SUM(ERR2)#/NOTEST)*100);
    * PRINT PR2 TERR2;

*THIS IS CLASSIFICATION METHOD 3 (BASED ON 1 HAZARD VALUE)*

DO J = 1 TO NOTEST;
    K = 1;
    DO K = 1 TO N1;
        IF H_T(J,) <= H1(K,) THEN GOTO OUT1;
    END;
    OUT1: ESTPR1(J,) = (N1-(K-1))#/N1;
END;
DO J = 1 TO NOTEST;
    K = 1;
    DO K = 1 TO N2;
        IF H_\(T(J)\) < H2(K,) THEN GOTO OUT2;
    END;
    OUT2: ESTPR2(J,) = (K-1)#/N2;
END;

DO L=1 TO NOTEST;
    IF ESTPR1(L,) > ESTPR2(L,) THEN CL3(L,) = 1;
    IF ESTPR1(L,) <= ESTPR2(L,) THEN CL3(L,) = 2;
    IF CL3(L,)=ACT_CL(L,) THEN ERR3(L,) = 0;
    ELSE ERR3(L,) = 1;
END;
* PR3 = NZ1 || NZ2 || NT || NIND || H_T || CL3 || ACT_CL:
    TERR3 = INT((SUM(ERR3)#/NOTEST)*100);
* PRINT PR3 TERR3;

*******************************************************************************;
*******************************************************************************;
* THIS IS FOR CLASSIFICATION METHOD 4 (COMPARE POST. PROB)******;
*******************************************************************************;

DO J = 1 TO NOTEST;
    IF P1_/T(J,) > P2_/T(J,) THEN CL4(J,) = 1;
    IF P1_/T(J,) <= P2_/T(J,) THEN CL4(J,) = 2;
    IF CL4(J,) = ACT_CL(J,) THEN ERR4(J,) = 0;
    ELSE ERR4(J,) = 1;
END;
* PR4 = NZ1 || NZ2 || NT || NIND || P1_/T || P2_/T || CL4 || ACT_CL:
    TERR4 = INT((SUM(ERR4)#/NOTEST)*100);
* PRINT PR4 TERR4;

*******************************************************************************;
*******************************************************************************;
* THIS IS CLASSIFICATION METHOD 5, ADAPTED FROM NEIGHBOR METHOD:
*******************************************************************************;

DO I=1 TO NOTEST;
    DO K=1 TO N1;
        DIFF1(K,) = H1(K,) - H_/T(I,);
        DIST1(K,) = DIFF1(K,) * INV(SP) * DIFF1(K,)
        END;
    NDIST1(I,) = MIN(DIST1);
END;

DO I=1 TO NOTEST;
    DO K=1 TO N2;
        DIFF2(K,) = H2(K,) - H_/T(I,);
        DIST2(K,) = DIFF2(K,) * INV(SP) * DIFF2(K,)
        END;
    NDIST2(I,) = MIN(DIST2);
END;
DIST = NDIST1 || NDIST2;

DO I = 1 TO NOTEST;
   IF DIST(I,1) < DIST(I,2) THEN CL5(I,1) = 1;
   IF DIST(I,1) >= DIST(I,2) THEN CL5(I,1) = 2;
   IF CL5(I,1) = ACT_CL(I,1) THEN ERR5(I,1) = 0;
   ELSE ERR5(I,1) = 1;
END;
* PR5 = NZ1 || NZ2 || NT || NIND || H_T || DIST || CL5 || ACT_CL;
   TERR5 = INT((SUM(ERR5)#/NOTEST)*100);
* PRINT PR5 TERR5;
*****************************************************************************;
*****************************************************************************;
* THIS IS THE NON-PARAMETRIC NEIGHBORHOOD CLASSIFICATION *
* PROCEDURE*;
*****************************************************************************;
MER1 = SUM(ER1)#/N1;
MER2 = SUM(ER2)#/N2;
MTD1 = SUM(TD1)#/N1;
MTD2 = SUM(TD2)#/N2;
* PRINT MER1 MER2 MTD1 MTD2;

S11 = (SUM((ER1' - MER1)*(ER1' - MER1')))}/#{N1 - 1};
S21 = (SUM((TD1' - MTD1)*(TD1' - MTD1')))}/#{N1 - 1};
COV1 = (SUM((ER1' - MER1)*(TD1' - MTD1')))}/#{N1 - 1};

S12 = (SUM((ER2' - MER2)*(ER2' - MER2')))}/#{N2 - 1};
S22 = (SUM((TD2' - MTD2)*(TD2' - MTD2')))}/#{N2 - 1};
COV2 = (SUM((ER2' - MER2)*(TD2' - MTD2')))}/#{N2 - 1};

T1 = (S11 || COV1) // (COV1 || S21);
T2 = (S12 || COV2) // (COV2 || S22);
S1P = ((N1 - 1)*S11 + (N2 - 1)*S12)}/#{N1+N2 - 2};
S2P = ((N1 - 1)*S21 + (N2 - 1)*S22)}/#{N1+N2 - 2};
COVP = ((N1 - 1)*COV1 + (N2 - 1)*COV2)}/#{N1+N2 - 2};
TP = (S1P || COVP) // (COVP || S2P);
TPINV = INV(TP);
*PRINT T1 T2 TP TINV;

DIST11 = J(N1,1,0);
DIST22 = J(N2,1,0);
NDIST11 = J(NOTEAST,1,0);
NDIST22 = J(NOTEAST,1,0);
**COMPUTE THE MINIMUM DISTANCE IN GROUP 1**;
DO I=1 TO NOTEST;
   DIFF11 = J(N1,1,0);
   DIFF21 = J(N1,1,0);
   DIS11 = J(N1,2,0);
   DO J=1 TO N1;
      DIFF11(J,) = ER1(J,) - NZ1(I,);
      DIFF21(J,) = TD1(J,) - NZ2(I,);
      DIS11(J,) = DIFF11(J,) || DIFF21(J,);
      DIST11(J,) = DIS11(J,) * TPINV * DIS11(J,);
   END;
   NDIST11(I,) = MIN(DIST11);
END;
* PRINT NDIST11;

**COMPUTE THE MINIMUM DISTANCE IN GROUP 2**;
DO I=1 TO NOTEST;
   DIFF12 = J(N2,1,0);
   DIFF22 = J(N2,1,0);
   DIS22 = J(N2,2,0);
   DO J=1 TO N2;
      DIFF12(J,) = ER2(J,) - NZ1(I,);
      DIFF22(J,) = TD2(J,) - NZ2(I,);
      DIS22(J,) = DIFF12(J,) || DIFF22(J,);
      DIST22(J,) = DIS22(J,) * TPINV * DIS22(J,);
   END;
   NDIST22(I,) = MIN(DIST22);
END;

DISTN = NDIST11 || NDIST22;
* PRINT DISTN;

**COMPUTE THE MINIMUM OF THE DISTANCES IN GR. 1 & GR. 2**;
DO I = 1 TO NOTEST;
   IF DISTN(I,1) < DISTN(I,2) THEN CLN(I,1) = 1;
   IF DISTN(I,1) >= DISTN(I,2) THEN CLN(I,2) = 2;
   IF CLN(I,1) = ACT_CL(I,1) THEN ERRN(I,) = 0;
   ELSE ERRN(I,) = 1;
END;
* FIN = NZ1 || NZ2 || DISTN || CLN || ACT_CL || ERRN;
   TERRN = INT((SUM(ERRN)##/NOTEST##100));
* PRINT FIN TERRN;

**PUT DIFFERENT ERROR RATES TOGETHER**;
ERROR = TERR1 || TERR2 || TERR3 || TERR4 || TERR5 || TERRN;
ERROR_2 = ERROR ## 2;
TERR(SIM,1) = TERR(SIM,1) + ERROR;
TERR_2(SIM,1) = TERR_2(SIM,1) + ERROR_2;
OUT: END;
**END OF SIMULATION **;

NNOSIM = NOSIM - FLAG;
NOCENS = SUM(CENS) #/ NNOSIM;
PRINT FLAG NOCENS;

MNPAR1 = SUM(FNPAR1) #/ NNOSIM;
MNPAR2 = SUM(FNPAR2) #/ NNOSIM;
MNPAR3 = SUM(FNPAR3) #/ NNOSIM;
MNPAR4 = SUM(FNPAR4) #/ NNOSIM;
MNPAR5 = SUM(FNPAR5) #/ NNOSIM;
MNPAR6 = SUM(FNPAR6) #/ NNOSIM;
FINPAR = MNPAR1 // MNPAR2 // MNPAR3 //
       MNPAR4 // MNPAR5 // MNPAR6 ;
PRINT FINPAR;

MVAR1 = SUM(FVAR1) #/ NNOSIM;
MVAR2 = SUM(FVAR2) #/ NNOSIM;
MVAR3 = SUM(FVAR3) #/ NNOSIM;
MVAR4 = SUM(FVAR4) #/ NNOSIM;
MVAR5 = SUM(FVAR5) #/ NNOSIM;
MVAR6 = SUM(FVAR6) #/ NNOSIM;
FINVAR = MVAR1 // MVAR2 // MVAR3 // MVAR4 // MVAR5 // MVAR6 :
PRINT FINVAR;

TEGR1 = SUM(TERR,(1));
TEGR2 = SUM(TERR,(2));
TEGR3 = SUM(TERR,(3));
TEGR4 = SUM(TERR,(4));
TEGR5 = SUM(TERR,(5));
TEGRN = SUM(TERR,(6));
TEGR1_2 = SUM(TERR_2,(1));
TEGR2_2 = SUM(TERR_2,(2));
TEGR3_2 = SUM(TERR_2,(3));
TEGR4_2 = SUM(TERR_2,(4));
TEGR5_2 = SUM(TERR_2,(5));
TEGRN_2 = SUM(TERR_2,(6));
MERR1 = TERR1/#/NNOSIM;
MERR2 = TERR2/#/NNOSIM;
MERR3 = TERR3/#/NNOSIM;
MERR4 = TERR4/#/NNOSIM;
MERR5 = TERR5/#/NNOSIM;
MERRN = TERRN/#/NNOSIM;
MEANERR = MERR1 || MERR2 || MERR3 || MERR4 || MERR5 || MERRN;

MINERR =>MIN(TERR,(1)) || MIN(TERR,(2)) || MIN(TERR,(3)) ||
       MIN(TERR,(4)) || MIN(TERR,(5)) || MIN(TERR,(6));
MAXERR =MAX(TERR,(1)) || MAX(TERR,(2)) || MAX(TERR,(3)) ||
       MAX(TERR,(4)) || MAX(TERR,(5)) || MAX(TERR,(6));
**END OF PROGRAM**
REFERENCES


Pollard, H. S. (1934). On the relative stability of the median and the arithmetic mean with particular reference to certain frequency distributions which can be dissected into normal distributions. The Annals of Mathematical Statistics. 5: 227-262.


