INFERENCE FOR RIGHT CENSORED, AND RIGHT CENSORED LENGTH BIASED DATA THROUGH INVERSE WEIGHTING

by

Chetachi Akunna Emeremni

B.S., StonyBrook University, 2004

M.A., Columbia University, 2006

Submitted to the Graduate Faculty of

the Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2012

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This dissertation was presented

by

Chetachi Akunna Emeremni

It was defended on

July 19, 2012

and approved by

Abdus S. Wahed, PhD.

Associate Professor of Biostatistics

Department of Biostatistics

Graduate School of Public Health

University of Pittsburgh.

Sati Mazumdar, PhD.

Professor of Biostatistics and Psychiatry

Department of Biostatistics

Graduate School of Public Health

University of Pittsburgh.

Jong-Hyeon Jeong, PhD.

Associate Professor of Biostatistics

Department of Biostatistics

Graduate School of Public Health

University of Pittsburgh.

Charity G. Moore, PhD.

Associate Professor of Biostatistics and Medicine

Center For Research on Health Care Data Center

University of Pittsburgh.

Dissertation Director: Abdus S. Wahed, PhD.

Associate Professor of Biostatistics

Department of Biostatistics

Graduate School of Public Health

University of Pittsburgh.

INFERENCE FOR RIGHT CENSORED, AND RIGHT CENSORED LENGTH BIASED DATA THROUGH INVERSE WEIGHTING

Chetachi Akunna Emeremni, PhD

University of Pittsburgh, 2012

In many medical studies, the outcome of interest may be the time from a starting point to a predefined specific event. A key feature of these data is that the complete event times may not be completely known for some subjects. When this occurs, the survival times are said to be censored. When observations are right censored, all that is known for such individuals is that their event time is greater than some given value. Analysis of variance has been one of the most powerful statistical tools for comparing mean continuous response across multiple groups. Use of classical ANOVA in time-to-event data is problematic because of the right censored nature

of survival times. In this dissertation, we propose a weighted analysis of variance approach to comparing mean continuous response between groups when the outcome is subject to right censoring. The method weights each observation by the inverse of the probability of being censored. We show that classical ANOVA methods such as decomposition of sums of squares and tests of contrasts follows in the weighted ANOVA setting. Simulation results show that the weighted ANOVA could be a comparable alternative to other methods of analyzing survival data. We apply our methods to a dataset from the North Central Cancer Treatment Group and another from the Radiation Therapy Oncology Group.

Length-Biased sampling is statistical artifact that occurs in survival analysis when the probability of an observation being included in the sample is proportional to a particular characteristic of that observation. When length-biased data is subject to right censoring, inference is biased if these key features of the data are not accounted for in the analysis. In this dissertation, we propose an estimating equation approach to eliminate the bias introduced by censoring and unequal sampling probability using inverse weighting.

Simulation results show that the estimator is a simple and effective method of analyzing survival data that is both right censored and length biased. The proposed method is applied to the Channing House data.

The mean survival time is often used as a measure of effectiveness in screening and other public health programs. When the data are right censored and/or length-biased, the mean survival time is a biased estimator of the population mean. This work is of important public health significance because it provides an effective method of comparing health populations using the mean survival time as the measure of effect.

Keywords: Analysis of variance, Survival analysis, Censoring, Inverse probability of censoring weighting, Length bias, Inverse probability weighting, Kaplan-Meier, Proportional hazard, Accelerated failure time, Buckley-James.

PREFACE

I would like to express my gratitude to my dissertation advisor Dr. Abdus S. Wahed who directed and guided me through the whole process of my study and research. I am greatly thankful to the members of the committee for their help and support in preparation of the dissertation. I also want to thank Dr. Sati Mazumdar for the support and training I received as a trainee under her NIMH T32 "Training Biostatisticians in Psychiatric Research" training grant. I would like to thank Dr. Jeong for his help, support and training to me as student in the department of Biostatistics. I would also like to acknowledge support from the the NIH RR024154 Multidisciplinary Clinical Research Program KL2 research program and thank Dr. Charity G. Moore for her support and mentorship while I served as a graduate student researcher under the program. I also want to express my sincere appreciation of the efforts of all the professors whose courses I have taken for they helped to create a solid foundation for my current and future research.

Finally I thank God for the grace to successfully complete this research and my family and friends for their encouragement, love and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

In many medical studies, the outcome of interest may be the time from a starting point to a specific predefined event. Examples of time to event data include time from diagnosis to death or remission, incubation time from HIV diagnosis to AIDS or time from birth to cessation of breastfeeding. Data arising from these studies with time-to-event outcomes are generally referred to as survival data. A key feature of these data is that the complete event times may not be completely known for some subjects. When this occurs, the survival times are said to be censored. Censoring occurs when the starting point is not observed (left censoring) or when the event has not yet occurred at the end of the observation period (right censoring). In both cases, the exact survival time cannot be fully determined. Right censoring occurs frequently in medical studies due to study termination or patient discontinuation. When observations are right censored, all that is known for such individuals is that their event time is greater than some given value. In the medical literature, the term censoring is often used to refer to 'right' censoring. In this paper, this convention is adopted as well.

Survival data arises from two data generating processes. One giving rise to the potential survival time T_j and another giving rise to a potential censoring time C_j for subject j. In this research, we assume that these processes are independent. What

is actually observed is $U_j = \min(T_j, C_j)$. Therefore the observed data consists of n independent and identically distributed copies of (U_j, δ_j), where δ_j is the indicator function which takes the value 1 if $T_j < C_j$. In analyzing survival data, two functions that are usually of particular interest are the survival and the hazard functions. The survival function $S(t) = \Pr(T_j > t$ is the probability of surviving beyond a specified time t while the hazard function $h(t)$ though not a probability can be thought of as the probability of failing in an infinitesimal time period beginning at time t given survival up to time t. Traditional methods of survival analysis typically rely on either of these two functions and include the Kaplan-Meier estimate, log-rank test, Cox proportional hazards model and the accelerated failure time models.

1.1 OBJECTIVE

The primary purpose of this dissertation is to improve upon and add to the existing methods for the analysis of right-censored, and right-censored length biased data. In particular, the methods presented in this dissertation extend the general linear analysis of variance model to right-censored data through the use of inverse probability weights. This method is further extended to the analysis of factorial designs with time to event outcomes. Inverse weighting methods are also applied to eliminate bias in the estimation of the mean survival time in right-censored length biased data.

We begin with a general overview of the most commonly used methods for the analysis of survival data and the benefits of hazards vs. time based modeling is discussed. In chapter 2, we present a general overview of the one way analysis of variance model and discuss challenges in extending the model to survival data and remedies for these challenges. In chapter 3, we discuss earlier works related to the

adaptation of the general linear model to survival data and discuss how the analysis of variance model can be extended to survival data. In chapter 4, we provide a general overview of inverse probability weighting and present details of the proposed method extending analysis of variance to right-censored survival data. In chapter 5, we provide details of the extension of the proposed method for the two-way analysis of variance model. In chapter 6, we discuss how inverse weighting techniques can be used to eliminate bias in the mean estimation of the survival time when data are both right-censored and length-biased. Finally in chapter 7, we provide a general discussion of the proposed methods and the general statistical and public health significance of the research.

1.2 THE KAPLAN-MEIER (KM) METHOD

The Kaplan-Meier method[25] is an empirical or non-parametric method of estimating the survival function for a given data. It is a popular method of analyzing survival data due to its simplicity and weak assumptions and is readily implemented in most statistical packages including R and SAS. The fundamental assumptions required for the KM test are (i) censoring is unrelated to prognosis (ii) survival probabilities are the same for all subjects irrespective of when they entered into the study and (iii) events happened at the times specified. The KM test also requires the usual independence of censoring times and survival times i.e. random dropout. Implementation of the KM method proceeds as follows:

1. Order the k observed survival times in increasing magnitude $0 \leq t_{(1)} \leq t_{(2)}.. \leq t_{(k)}$. When a censored observation has the same time as uncensored observation,

the common convention is that uncensored value should precede the censored value in the ordered list.

2. Let t_j be the j^{th} unique value in the series of ordered survival times, r_j be the number of subjects at risk at time until just before time t_j, and d_j be the number of subjects who fail at time t_j .

3. Then for each j calculate,

$$\hat{S}(t_j) = \prod_{j=1}^{n} \frac{r_j - d_j}{r_j}$$

$\hat{S}(t_j)$ is the estimate of the survival function at time t_j. The KM estimate is a consistent estimate of the survival function within the range of times in the given data. Modifications to the KM method to evaluate $\hat{S}(t_j)$ for t greater than the maximum value t_{max} of observed survival times have been proposed by : Efron [14] who suggested $\hat{S}(t) = 0$ for $t \geq t_{max}$, Gill [16] who suggested $\hat{S}(t) = \hat{S}(t_{max})$ and Brown et al. [5] who suggested $\hat{S}(t) = exp[log(\hat{S}(t_{max}))t/t_{max}]$. However, the validity of $\hat{S}(t)$ beyond the range of the data cannot be assessed. The variance of the KM estimator of the survival function developed by Greenwood [19] is given by $\hat{V}(\hat{S}(t_j)) = \hat{S}(t_j)^2 \sum_{j=1}^{n} \frac{d_j}{(r_j)(r_j - d_j)}$.

1.3 THE LOG RANK TEST

The log rank (LR) test [34] is a statistical hypothesis test for comparing two survival curves. The LR tests the hypothesis of no difference in the population survival curves of two groups. Intuitively, the LR tests whether the probability of any event at any time point is the same between the two groups. The LR test is conducted by

constructing a $2x2$ table at each distinct death time and comparing the observed and expected death rates between the two groups conditional on the number of subjects at risk at time point. The test statistic for the LR test is then given by:

$$\sum_{j}^{n} \frac{(O_j - E_j)^2}{E_j} \sim \chi_1^2 \quad \text{under the null}$$

where O_j and E_j are the observed and expected number of events in group i assuming no difference in the probability of an event between the two groups. The assumptions for the LR test are the same as those required for the KM test. Deviations from these assumptions can affect the inference from the test when they are satisfied differentially across the groups being compared. The LR test is most likely to detect a difference in the survival curve of one group consistently lies above the other. It is unlikely to detect a difference in the curves when the survival curves of the two groups cross. The LR test is most powerful under a proportional hazards assumption. The log-rank test is merely a test of significance and does not quantify the degree of the difference between the groups. It cannot be used when other covariates associated with survival need to be adjusted for.

1.4 COX SEMI PARAMETRIC PROPORTIONAL HAZARDS MODEL

The proportional hazards (PH) model [9] models the effect of covariates on the hazard function. The PH model is given by $h_j(t) = h_0(t) * exp(\beta_1 x_{j1} + \cdots + \beta_p x_{jp})$. The predictors $X_1, X_2, .., X_p$ are assumed to act additively on the log hazard i.e the $logh(t|X)$ changes linearly with X. The model is semi parametric because the

functional form of the baseline hazard does not need to be specified. Cox's PH model assumes that the ratio of the hazards for any two individuals is independent of time. Consider two subjects l and j, the ratio of their hazards at any given time is

$$\frac{h_l(t) = h_0(t) * exp(\beta_1 x_{l1} + \cdots + \beta_p x_{lp})}{h_j(t) = h_0(t) * exp(\beta_1 x_{j1} + \cdots + \beta_p x_{jp})}.$$

Note that the baseline hazard cancels out and the ratio is therefore a function of only observed covariates x i.e the effect of the predictors is the same at all time points. The assumption of proportional hazards is beneficial in the estimation of the regression parameters as it greatly simplifies the likelihood function. If there are d subjects at risk at time k, assuming proportional hazards, the probability that it is subject l who fails at time k is given by

$$\begin{aligned} L_l &= \Pr(\text{subject l fails at time k}) \\ &= \frac{h_l(t)}{h_1(t) + h_2(t) + \cdots + h_d(t)} \\ &= \frac{exp(\boldsymbol{\beta X_l})}{\sum_{j \in R_k} exp(\boldsymbol{\beta X_j})} \end{aligned}$$

where R_k is the risk set at time k. L_l is subject l's contribution to the partial likelihood. Thus the partial likelihood function is:

$$l(\beta) = \prod_{t_k} \frac{exp\beta x_i}{\sum_{j \epsilon R_k} exp(\beta x_j)}.$$

Cox [9] and others have shown that this partial likelihood can be treated as an ordinary likelihood to derive valid estimates for the regression parameters. Therefore estimation of the $\boldsymbol{\beta}$ proceeds using standard maximum likelihood theory. This partial likelihood function however is valid only when there are no ties present in the data. In the presence of ties, the true partial log likelihood function is difficult to permute

and the Efron [15] and Breslow [4] approximation methods to the likelihood function can be used.

1.5 THE ACCELERATED FAILURE TIME MODEL

The accelerated failure time (AFT)[56] model is useful in assessing the direct effect of covariates on the survival rather than on the hazard. The AFT model is given by:

$$log(T) = X\beta + \epsilon \quad \text{where } \epsilon \text{ has a distribution function F}$$

This is equivalent to $T = exp(X^T\beta)v$ where $v = exp(\epsilon)$. That is, the effect of covariate X is to act multiplicatively on survival time. The contribution to the likelihood function at time t_j for a subject who failed is the density function $f(t_j;\theta;x_j)$ because the exact lifetime is know while the contribution is the survival function $S(t_j;\theta;x_j) = 1 - F(t_j;\theta;x_j)$ for an observation who is right censored because all that is known is that the true event time is larger than t_j. Therefore the likelihood function for right-censored survival data is proportional to

$$\text{Ł}(\theta) \propto \prod f(t_j;\theta;x_j)^{\delta_j} F(t_j;\theta;x_j)^{1-\delta_j}$$

where δ_j is an indicator function for censoring taking the value 1 for observations that are uncensored. In the parametric AFT model, all components are completely specified. Parameter estimation of model parameters are obtained using the method of maximum likelihood. AFT models have the advantage of ease of interpretation because covariates are acting directly on the mean survival time.

1.6 HAZARD VS. TIME BASED MODELING

The Cox proportional hazards model has become the model of choice for the analysis of time to event data. In the Cox model, β is the estimate of the treatment effect. The hazard ratio is a ratio of the hazard functions between the 'treated' group and the 'control' group and is given by $exp(\beta)$. The hazard ratio is only useful when it has a meaningful interpretation. In clinical practice, physicians and their patients are generally interested in the benefit of treatment in prolonging survival. However, a relative reduction in hazards does not directly reflect an equivalent relative improvement in survival. The ratio of survival probabilities in two groups at any given time typically will not equal the hazard ratio. A fundamental assumption of the Cox model is that of proportional hazards i.e the hazard ratio is constant over time. This assumption may not be valid in certain cases. The violation of the proportional hazards assumption casts doubt on the use of the hazard ratio as a measure of treatment effect.

In classical regression models of quantitative outcomes, effects of covariates on the outcome are modeled via the mean. Although the restricted mean survival time for any treatment group i can be calculated from the fit of a Cox model using $S_i(t) = S_0(t)^{exp(\hat{\beta}_i)}$ where $S_0(t)$ is the baseline survival function. The validity of the Cox model is brought into question when the PH assumption is violated. When the proportionality of hazard rates is not maintained over time, it leads to biased parameter estimates since the method of estimation is based on this proportionality. The AFT model directly models the effect of covariates on mean survival time. However, although AFT models are readily implemented in standard available software packages, it is not commonly used for the analysis of clinical trials. Comparison of mean survival times between groups is an important goal in many clinical applications.

For example, the National Institute on Aging developed an intervention testing program RFA-AG-02-005 [38] whose primary outcome objective was to increase mean life expectancy by 10% in phase 1 studies. Other examples include studies in health economics where the primary measure of effectiveness includes the mean survival time. When the mean survival time is used as the measure of effectiveness, then the calculated incremental cost effectiveness ratio has the simple interpretation as the cost of achieving an extra year of life.

In the absence of censoring, survival times are completely observed for all patients and the sample mean and variance are consistent estimators for the population mean and variance survival time. However, in the presence of right censoring, the usual estimate of the mean survival time is not appropriate. The mean survival time is given by $\int_0^\infty S(t)\mathrm{d}t$. Because of censoring, the entire survival distribution is not observable. Irwin [22] defined instead the restricted mean survival time. The restricted mean survival (rms)time of a random variable T measured from a predefined starting point such as diagnosis or study entry up to a given time τ is defined as the $\min(T, \tau)$ and is given by $\int_0^\tau S(t)\mathrm{d}t$. The (unrestricted) mean survival time is therefore the limit of $\mu(\tau)$ as $\tau \sim \infty$. Restricted mean survival time can be estimated non-parametrically using the Kaplan Meier estimator for the survival function.

Direct regression analysis of the mean (log) survival time is accomplished through the accelerated failure time model. This is a standard linear regression model for the log survival time that accounts for the right censored nature of the data. Estimation of the mean parameter can be accomplished through maximum likelihood estimation when a parametric model is specified for the distribution of the residuals. Semi parametric estimation procedures when the residual distribution is left unspecified were also developed by Buckley and James[6]. Miller [36] introduced en estimation procedure for the regression parameters in a linear model when the data are randomly

right censored based on the method of least squares and Koul Susarla and Van Ryzin [28] proposed a method for handling linear regression models with censored data based on a re-weighting technique.

2.0 ANALYSIS OF VARIANCE

Analysis of variance (ANOVA) has been one of the most powerful statistical tools for comparing means of continuous response across multiple groups. ANOVA is easy to understand and interpret to people with limited statistical knowledge, as it is based on decomposing the total variation in the data into components due to known (e.g. group) and unknown (e.g. random error) factors. Inference from the linear model is done through a comparison of the component variations to that of the error variation. If the variation due to one factor is significantly larger than the variation due to error, that factor is deemed to be significantly associated with the response variable.

2.1 THE ONE-WAY ANOVA MODEL

Consider the classical one-way ANOVA model

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij} = \mu_i + \epsilon_{ij}, \quad i = 1, 2, ..., p; j = 1, 2, ..., n_i \tag{2.1}$$

where Y_{ij} is the response variable measured on the j^{th} subject in the i^{th} treatment group, μ is the intercept term (grand mean), α_i is the treatment effect for the i^{th}

treatment, ϵ_{ij} is the random error term associated with the $(i,j)^{th}$ subject, p is the number of levels of the treatment and n_i is the number of subjects in group i. The second representation of the one-way ANOVA model is called the cell means model. Here μ_i has the interpretation as the means corresponding to the levels of the treatment. Assumptions for the model (2.1) are:

1. $E(\epsilon_{ij}) = 0; \quad Var(\epsilon_{ij}) < \infty \quad$ for all i,j

2. ϵ_{ij}'s are independently and identically distributed as $N(0, \sigma^2)$.

The one-way ANOVA model tests the hypothesis $H_0 = \mu_1 = \mu_2 = \cdots = \mu_p$. The associated test statistic is based on examining the total variability in the data. If the null hypothesis is true then we would expect the variation in the data to be due to chance. Otherwise, we expect that it is due to differences among the treatment groups. The variability in the data is expressed as a sum of squared deviations. This total variability (SST) can be decomposed into a contribution due to treatment groups (SSTR) and a contribution due to error (SSE) as follows:

$$SST = SSTR + SSE$$

$$\sum_i \sum_j (Y_{ij} - \bar{Y}_{..})^2 = \sum_i \sum_j (\bar{Y}_{i.} - \bar{Y}_{..})^2 + \sum_i \sum_j (Y_{ij} - \bar{Y}_{i.})^2.$$

Then using the normality assumption, $SSTR/\sigma^2$ and SSE/σ^2 are independent χ^2 random variables with $(p-1)$ and $(N-p)$ degrees of freedom respectively, where p is the number of treatment group levels and N is the total number of subjects. Since the expected value of χ^2 random variable is equal to its degrees of freedom, the mean square error (MSE)=$\frac{SSE}{(N-p)}$ and the treatment mean square (MSTR)=$\frac{SSTR}{(p-1)}$ are estimates of the variance σ^2 under the null. Therefore, the test statistic for testing

Table 2.1: One-way ANOVA Table

Source	df	Sum of Squares (SS)	Mean Squares (MS)	F
Treatment	p-1	SSTR=$\sum_i \sum_j (\bar{Y}_i - \bar{Y}_{..})^2$	MSTR=$\frac{SSTR}{p-1}$	
Error	n-p	SSE=$\sum_i \sum_j (Y_{ij} - \bar{Y}_{i.})^2$	MSE=$\frac{SSE}{n-p}$	$\frac{MSTR}{MSE}$
Corrected Total	n-1	TSS=$\sum_i \sum_j (Y_{ij} - \bar{Y}_{..})^2$		

the alternative hypothesis is given as:

$$F = \frac{SSTR/((p-1))}{SSE/((N-p))} = \frac{MSTR}{MSE} \quad \sim F_{(p-1),(N-p)} \quad \text{under the null}$$

The data from an experiment employing one-way ANOVA is summarized in an ANOVA table as shown in Table (2.1).

2.2 OLS FOR SURVIVAL DATA

The challenges with using OLS for survival analysis arises from the nature of survival times. In ordinary least squares regression, for inferential reasons, the distribution of the error terms ϵ_{ij} are assumed to follow a normal distribution. This implies that the outcomes conditional on the covariate vector also follow a normal distribution. This assumption is often unreasonable for survival times since distributions of survival time may have gross departures from normality. Survival times are positive, often right skewed random variables. In contrast, the normal distribution has zero skewness

and is symmetric about its' mean. While OLS regression may be robust to slight departures from normality, it is not robust when the underlying distribution is non-symmetrical. The inference from an analysis depends upon the assumptions being made and when these assumptions fail, the inference is rendered invalid.

Another challenge with OLS for survival data is that survival times may not be completely known exactly due to right censoring. Since under censoring we observe only the minimum of the failure and censoring time, the usual linear model in 2.1 does not hold and therefore the least squares approach is not applicable.

2.3 EXPLAINED VARIATION

Pearson correlation coefficients and other similar nonparametric statistics are important tools in data analysis to quantify the magnitude of the relationship between two continuous variables. The amount of explained variation is one way of quantifying the degree to which a relationship is present. The coefficient of determination (R^2) is the proportion of variation in the response explained by the regression equation and is equal to the square of the correlation between the response and a linear combination of the covariates. In the one way ANOVA model, R^2 is defined as $1 - \frac{errorSS}{totalSS}$. That is, R^2 is the proportion of the total variation that is explained by group membership. When $R^2 = 1$, all of the variation is explained by the treatment groups, when $R^2 = 0$, none of the variation is explained by group membership. In survival analysis, the sum of squared deviations cannot be completely determined because of censoring. Several adaptations of R^2 have been established in survival. We discuss the statistic proposed by Nagelkerke [37] which is output by standard software packages such as the *psm* function in the R Design library. This statistic is given

by $R^2 = 1 - exp(\frac{LR^2}{n})$ where LR is the likelihood ratio chi square statistic for testing that all coefficients in the model is equal to zero. This definition does not have the literal meaning as the proportion of variance explained as in the least squares coefficient of determination. However, it is a measure of how strongly associated the independent variables are with the dependent variable. The ratio is indicative of the degree of improvement in prediction from the null model. It is still questionable as to what should be the equivalent of R^2 for survival analysis. However, according to the guidelines postulated in Shemper and Stare [45], properties of a good R^2 measure include (i) the expected value of the measure should not be affected by censoring, although censoring may limit the precision with which the measure is estimated, (ii) it should not be affected by monotonic transformations of the data and (ii) should have a clear and intuitive interpretation.

In the subsequent chapters, we discuss an analysis of variance type analysis for survival data in analogy to the analysis of variance for the classical linear model. An ANOVA decomposition is established and a coefficient of determination R^2 is established to describe the effect of group membership . This measure of R^2 has an intuitive explanation as the proportion of variation explained as in the classical least squares coefficient of determination. An ANOVA-like F test is established for testing the null hypothesis of no effect of treatment on mean survival time. We investigate the performance of estimators for the mean survival time using Monte Carlo simulations. We will compare the results of the comparison of mean survival times between treatment groups using the proposed method with the nonparametric Kaplan Meier estimator and the accelerated failure time model. The usefulness of the proposed method will be illustrated using a real data set from the Radiation Therapy Oncology Group [24].

3.0 LINEAR REGRESSION FOR CENSORED DATA

Because of censoring many statistical tools including analysis of variance can not be used in their standard form and need to be adapted to accommodate the fact that some observations are not fully observed. The ANOVA model is a simple special case of the linear model. The linear regression model that relates covariates linearly to the survival time is an appealing one due to its ease of interpretation. The censored linear regression model has been the subject of much statistical research. Notable contributors to this field of study based on the method of least squares include the work by Miller [36], Buckley and James [6] and Koul, Susarla and Van Ryzin [28].

Consider the standard linear regression model:

$$T_j = \alpha + \beta x_j + \epsilon_j; \quad \epsilon_j \sim F; \quad E(\epsilon_j) = 0; \quad Var(\epsilon_j) = \sigma^2 \tag{3.1}$$

In the uncensored data setting, estimates of α and β are those which minimize $\sum (T_j - \alpha - \beta x_j)^2$ which is equivalent to minimizing the function:

$$\frac{1}{n} \sum_{j}^{n} (T_j - \alpha - \beta x_j)^2 = \int z^2 d\hat{F}_{\alpha,\beta}(z),$$

where $z = (T_j - \alpha - \beta x_j)$ and $\hat{F}_{\alpha,\beta}$ is the usual empirical distribution function of the error term. Miller reasoned that when the data is subject to right censoring, this is

equivalent to minimizing $\int z^2 d\hat{F}^{KM}_{\alpha,\beta}(z)$ where $\hat{F}^{KM}_{\alpha,\beta}$ is the non-parametric Kaplan-Meier estimator of the distribution function. For fixed values of α and β, this is the weighted sum of squares over uncensored values where the weights are equal to the size of the jumps assigned to the $z_j's$ by the Kaplan-Meier estimator. Estimates of β are obtained through an iterative procedure. Then for fixed β, an estimate of α can be easily obtained. A sufficient condition which must be imposed to maintain the asymptotic consistency of the estimates is that as x changes, the censoring distribution shifts along the same regression line as the survival distribution.

To overcome the consistency problems of Miller's method, Buckley and James introduced a modification of the usual least squares normal equations to account for the right censored survival times. Note that when a response variable is censored, the observed response can be written as: $U_j = T_j\delta_j + C_j(1-\delta_j)$ where T_j is the potential survival time and C_j is the potential censoring time, $U_j = min(T_j, C_j)$ is the observed response and δ_j is the indicator function which takes the value 1 when the observation is uncensored. Buckley and James noted that when the censored observations are replaced by their 'true' conditional values, the linearity in the observed responses is maintained i.e $U_j^* = T_j\delta_j + E(T_j|T_j > C_j)(1-\delta_j)$. This is shown by conditioning on δ_j as follows:

$$\begin{aligned}
E(U_j^*) &= E(U_j^*|\delta_j = 1)\Pr(\delta_j = 1) + E(U_j^*|\delta_j = 0)\Pr(\delta_j = 0) \\
&= E(T_j|\delta_j = 1)\Pr(\delta_j = 1) + E(E(T_j|T_j > C_j)|\delta_j = 0)\Pr(\delta_j = 0) \\
&= E(T_j|\delta_j = 1)\Pr(\delta_j = 1) + E(T_j|T_j > C_j)\Pr(\delta_j = 0) \\
&= E(T_j|\delta_j = 1)\Pr(\delta_j = 1) + E(T_j|\delta_j = 0)\Pr(\delta_j = 0) \\
&= E(T_j) = \alpha + \beta x_j.
\end{aligned}$$

Estimation of the regression parameters would be straightforward if $E(T_j|T_j > C_j)$ were known. By initially setting $\alpha = 0$, the conditional expected values are estimated from the data. Then α and β are estimated iteratively. The iteration proceeds first with the estimation of the slope parameter and later with the intercept parameter. The Buckley and James method differs from that of Miller in that the normal equations rather than the sum of squared residuals are modified in the censored data setting. The properties of the Buckley-James estimator have been studied extensively by many authors. James and Smith[23] studied the consistency of the Buckley-James estimator under regularity conditions and avoiding assumptions of the censoring patterns. The asymptotic properties of the Buckley-James estimator have been investigated by Lai and Ying[31]. They showed that with slight modifications, any consistent solution to the Buckley-James estimating equation is asymptotically normal and is semi-parametrically efficient when the normal distribution is assumed as the underlying error distribution.

Koul, Susarla and Van Ryzin [28] proposed a new estimator for the randomly right censored data when the error distribution is unknown. They noticed that under the assumption that survival times follow the model (3.1) where the errors are independent and identically distributed (i.i.d) with mean zero and independent of the censoring times, $E[\delta_j T_j G(T_j)^{-1}] = E[T_j]$, where $G(t) = P(C_j > t)$ is the survival function of the censoring time C_j. Hence the variables $Z_j = \delta_j T_j / G(T_j)$ have the same mean as T_j and thus follows the same linear model as T_j except with new errors that may no longer be identically distributed.

Note that $\delta_j T_j / G(T_j)$ is not observable because G is not a known function. Koul, Susarla and Van Ryzin propose estimating G by

$$\hat{G} = \prod_{j=1}^{n} \{(1 + N^+(U_j))/(2 + N^+(U_j))\}^{[\delta_j=0, U_j \leq t]} \quad -\infty \leq t \leq \infty,$$

where $N^+(y)$ is the number of U_j exceeding y. An advantage of their method is that it does not require an iterative procedure to obtain the estimates. The estimators $\hat{\alpha}$ and $\hat{\beta}$ derived are shown to be consistent and asymptotically normally distributed. Miller and Halpern [35] evaluated the performances of the Miller, Buckley-James and Koul et al estimators and concluded that the Buckley-James estimator is more reliable than the estimators by both Miller, and Koul et al.

3.1 THE ANOVA MODEL FOR SURVIVAL DATA

Let T_{ij} denote the survival time and C_{ij} denote the potential censoring time for the j^{th} patient in the i^{th} group. Note that survival times are not observed for subjects who are right censored. We only observe (U_{ij}, δ_{ij}) where $U_{ij} = min(T_{ij}, C_{ij})$ and $\delta_{ij} = 1$ if $T_{ij} \leq C_{ij}$ and 0 otherwise. We postulate the standard analysis of variance model for the log transformed survival times as

$$Y_{ij} = ln(T_{ij}) = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1, 2, ..., p; j = 1, 2, ..., n_i, \tag{3.2}$$

where μ is the intercept, α_i is the effect of the i^{th} group level, ϵ_{ij} is the error term associated with person j in group i and $\epsilon'_{ij}s$ are iid $N(0, \sigma^2)$. A log transformation was used in this model to reduce the skewness of the survival data.

For simplicity, let us define $\mu_i = \mu + \alpha_i$. Here μ_i has the interpretation of the

mean of the log survival for the i^{th} group. The hypothesis to be tested in (3.2) is: $H_0 : \alpha_1 = ... = \alpha_p$ or equivalently $\mu_1 = \mu_2 = ... = \mu_p$. A defining property of classical ANOVA is that it provides a useful way of thinking about how variation in the data can be attributed to different sources. When observations are complete, the total variation in the data, expressed as the sum of the squared deviations about the mean can be decomposed into two key components: the between and within variations as follows

$$\sum_i \sum_j (Y_{ij} - \bar{Y}_{..})^2 = \sum_i \sum_j (\bar{Y}_{i.} - \bar{Y}_{..})^2 + \sum_i \sum_j (Y_{ij} - \bar{Y}_{i.})^2. \tag{3.3}$$

When data are subject to right censoring, the above methods of sum of squares decomposition and the F-test fail because failure times are unknown for censored observations and therefore the squared deviation cannot be computed for these outcomes. When interest is in estimating the regression coefficients of the response on the explanatory covariates and summarizing the analysis in terms of the decomposition of the sums of squares for each source of variation in the model, methods that appropriately handle censoring are essential. In the following chapter, we discuss an adaptation of the traditional analysis of variance and method of least squares for right censored survival data under the assumption that log-survival times follow the model (3.2).

4.0 INVERSE PROBABILITY OF CENSORING WEIGHTED ANOVA OR IPCW ANOVA

Suppose we wish to apply the method of least squares or its variation to right censored survival data to fit the model (3.2) and explain the total variability in the data as a sum of a component due to the factor and a component due to chance variation. In the presence of censoring, the failure times are unknown when $\delta_{ij} = 0$, which can be viewed as a missing data problem. A general approach for handling missing data when data are missing at random is based on inverse probability weighting[32] which is a generalization of the Horvitz-Thompson estimator [20]. In the presence of censoring, observations whose failure times are observed are weighted by their inverse probability of being observed at their event times. This results in the creation of a pseudo-population; one we would have observed if there were no censored observation.

4.1 INVERSE PROBABILITY WEIGHTING

We consider methods for estimating a population mean when data on the response or dependent variable is missing. If we can assume that the data are missing com-

pletely at random (MCAR) i.e the probability that a case is missing is independent of the unobserved outcomes $\boldsymbol{Y}$ and observed covariates $\boldsymbol{X}$, then the simplest and easiest approach is to ignore it and proceed with a complete-case (CC) analysis. CC analysis focuses only on observations with no missing values and the CC estimator is a consistent estimator because when data are MCAR, the sample is still representative of the original population. Alternatively, when the probability of a case being missing is independent of $\boldsymbol{Y}$ given $\boldsymbol{X}$ i.e $\Pr(R = r|X, Y) = \Pr(R = r|X)$ where $\boldsymbol{R}$ is an indicator function taking the value of 1 for units that are observed. The data is said to be missing at random (MAR) and an analysis based on complete cases alone will be biased and inefficient.

Inverse probability weighting (IPW) is a modification of the complete case analysis. Inverse probability weighting was first introduced in the context of survey sampling to correct for unequal sampling sampling fractions [20]. In the IPW approach, the IPW estimator $\hat{\theta}_{IPW}$ is the solution of the IPW estimating equation equation $\sum_j^n \frac{R_j U_j(\theta)}{\pi_j} = 0$ where $U_j(\theta)$ is the function for estimating θ if all values of the response had been completely observed and π_j is the positive weight given to the j^{th} individual which is equal to the inverse of their probability of being a complete case. Since π_j is the probability of being a complete case, the IPW estimating equation has expectation zero and therefore parameter estimates are consistent. Also, each complete case represents π_j^{-1} units in the population with incomplete data.

When the missingness is due to censoring, then the probability of a complete case is $\Pr(C_j > T_j)$. Then the inverse of probability of censoring weight is equal to $\{\Pr(C_j > T_j)\}^{-1}$.Typically $\Pr(C_j > T_j)$ may not be known and needs to be estimated from the data. This is done using the Kaplan-Meier estimator. Use of the Kaplan-Meier estimator is based on the assumption that censoring times are random and independent of the survival times. However, inverse probability of censoring

weights are still valid even when the censoring process can be modeled using known variables. In this case, the Kaplan-Meier estimator can no longer be used as the basis for estimating the weights. Rather, regression models such as the Cox-PH can be used to estimate the survival probabilities of the censoring times. Then as with any regression model, variables should be included with caution because as the number of included variables increases, some incomplete cases may have a zero predicted probability assigned to them because the model perfectly predicts the data.

5.0 THE IPCW ANOVA MODEL

The model under consideration is the model 3.2, namely,

$$Y_{ij} = ln(T_{ij}) = \mu + \alpha_i + \epsilon_{ij} = \mu_i + \epsilon_{ij}$$

where μ_i is the population mean in treatment group i and ϵ_{ij} is the random error term associated with the j^{th} subject in the i^{th} treatment group.

In the least squares method, the sum of the squared deviations of each observation from its theoretical mean is minimized with respect to the model parameters. Thus, in the ordinary ANOVA model, one seeks to minimize

$$Q = \sum_{i}^{p} \sum_{j}^{n_i} (Y_{ij} - \mu_i)^2 = \sum_{i}^{p} \sum_{j}^{n_i} (lnT_{ij} - \mu_i)^2. \tag{5.1}$$

For a given μ_i, the quantity Q cannot be computed from the data, for some observations are censored. Now suppose $K_i(y) = P(lnC_{ij} > y)$ denotes the survival distribution of the lnC_{ij} in the i^{th} group. Then using the method of inverse probability weighting [42], a person whose survival time is completely observed at time T_{ij} has a probability $K_i(ln(T_{ij})) = K_i(Y_{ij})$ of not being censored, thus such a person whose survival time is completely observed is on average representative of $1/K_i(Y_{ij})$ similar censored individuals in the i^{th} treatment group.

In the presence of censoring, this suggests minimizing instead the weighted function

$$Q^* = \sum_i^p \sum_j^{n_i} \frac{\delta_{ij}}{K_i(U_{ij})}(U_{ij} - \mu_i)^2,$$

which is the standard objective function Q in 5.1 except that each uncensored observation is weighted inversely by the probability that the corresponding event is a failure. Note that if $K_i(.)$ is known then Q^* can be computed from the data for a given μ_i, and $\delta_{ij}h(Y_{ij}) = \delta_{ij}h(U_{ij})$ for all $h(.)$. Moreover,

$$\begin{aligned} E(Q^*) &= E\left[\sum_i \sum_j \frac{\delta_{ij}}{K_i(Y_{ij})}(Y_{ij} - \mu_i)^2\right] \\ &= \sum_i \sum_j E\left(E\left[\frac{\delta_{ij}}{K_i(Y_{ij})}(Y_{ij} - \mu_i)^2 \middle| Y_{ij}\right]\right) \\ &= \sum_i \sum_j E\left[\frac{(Y_{ij} - \mu_i)^2}{K_i(Y_{ij})}\right] E(T_{ij} < C_{ij}|Y_{ij}). \end{aligned}$$

But

$$E[T_{ij} < C_{ij}|Y_{ij}] = E[lnT_{ij} < lnC_{ij}] = P(lnC_{ij} > Y_{ij}) = K_i(Y_{ij}), \qquad (5.2)$$

The first equality in 5.2 follows, since the censoring is independent of survival time and the natural logarithm is a monotone function. Therefore,

$$\begin{aligned} E[Q^*] &= \sum_i \sum_j E\left(\frac{(Y_{ij} - \mu_i)^2}{K_i(Y_{ij})} K_i(Y_{ij})\right) \\ &= E\left[\sum_i \sum_j (Y_{ij} - \mu_i)^2\right] \\ &= E[Q]. \end{aligned}$$

Thus, minimization of Q^* is equivalent to minimizing Q in large samples.

When $K_i(.)$ is known, minimization of Q^* with respect to μ_i leads to the weighted least square estimator:

$$\hat{\mu}_i = \sum_j^{n_i} \frac{\delta_{ij} U_{ij}}{K_i(U_{ij})} / \sum_j^{n_i} \frac{\delta_{ij}}{K_i(U_{ij})} = \frac{\sum_j^{n_i} W_{ij} U_{ij}}{\sum_j^{n_i} W_{ij}} = \frac{\sum_n^{n_i} W_{ij} U_{ij}}{W_{i.}},$$

where $W_{ij} = \frac{\delta_{ij}}{K_i(U_{ij})}$, and $W_{i.} = \sum_j^{n_i} W_{ij}$.

Lemma 1: When $K_i(.)$ is known, the following results hold:

(i) $E(W_{ij}) = 1$

(ii) $E[W_{ij} h(U_{ij})] = E[h(Y_{ij})]$ for any function $h(.)$.

Proof:

(i)
$$\begin{aligned} E(W_{ij}) &= E\left[\frac{\delta_{ij}}{K_i(U_{ij})}\right] = E\left[\frac{\delta_{ij}}{K_i(Y_{ij})}\right] \\ &= E\left[E\left(\frac{\delta_{ij}}{K_i(Y_{ij})}\middle| Y_{ij}\right)\right] \\ &= E\left[\frac{1}{K_i(Y_{ij})} E(\delta_{ij}|Y_{ij})\right] \\ &= E\left[\frac{K_i(Y_{ij})}{K_i(Y_{ij})}\right] = 1. \quad \textit{by (5.2)} \end{aligned}$$

(ii)
$$\begin{aligned} E[W_{ij} h(U_{ij})] &= E\left[\frac{\delta_{ij} h(Y_{ij})}{K_i(Y_{ij})}\right] \\ &= E\left(E\left[\frac{\delta_{ij} h(Y_{ij})}{K_i(Y_{ij})}\middle| Y_{ij}\right]\right) \\ &= E\left[\frac{h(Y_{ij})}{K_i(Y_{ij})} E(\delta_{ij}|Y_{ij})\right] \\ &= E\left[\frac{h(Y_{ij})}{K_i(Y_{ij})} K_i(Y_{ij})\right] \\ &= E[h(Y_{ij})]. \end{aligned}$$

The consistency of $\hat{\mu}_i$ follows from part (ii) of Lemma 1 with $h(y) = y$. It is shown later that when $K_i(.)$ is known, $\hat{\mu}_i$ is asymptotically normally distributed.

Since $K_i(.)$ is unknown, it is estimated by the Kaplan-Meier estimator $\hat{K}_i(.)$ of the censoring times within the i^{th} group

$$\hat{K}_i(t) = \prod_{j:lnU_{ij} \leq t} \left(1 - \frac{1 - \delta_{ij}}{R(U_{ij})}\right), \tag{5.3}$$

where $R(U_{ij})$ is the number at risk of failing at time U_{ij}. Since within each group, $W_{ij}Y'_{ij}s$ are independently distributed for known $K_i(.)$, $\frac{1}{n_i}\sum_j^{n_i} W_{ij}Y_{ij}$ is asymptotically distributed as normal with mean μ_i and variance $\sigma^2_{\hat{\mu}_i}$, to be derived later. Therefore,

$$\hat{\mu}_i = \sum_j^{n_i} W_{ij}Y_{ij} / \sum_j^{n_i} W_{ij} \tag{5.4}$$

$$= (\bar{W}_{i.})^{-1} \frac{\sum_j^{n_i} W_{ij}Y_{ij}}{n_i}, \tag{5.5}$$

where $\bar{W}_{i.}$ is the sample group mean of weights.

By the first result in Lemma 1 and by the weak law of large numbers, $\bar{W}_{i.} \xrightarrow{p} 1$. Applying Slutsky's theorem to equation (5.4), $\hat{\mu}_i \xrightarrow{d} N(\mu_i, \sigma^2_{\hat{\mu}_i})$. The asymptotic properties of the weighted estimator can be derived by using the M-estimator theory [21][51]. For $K_i(.)$ known, the estimator $\hat{\mu}_i$ is a solution to the equation $\sum_j^{n_i} \psi_{ij}(U_{ij}, \delta_{ij}; \mu_i) = 0$, where $\psi_{ij}(U_{ij}, \delta_{ij}; \mu_i) = W_{ij}(U_{ij} - \mu_i)$ Let us define $A(\mu_i) = -E\left[\frac{\partial}{\partial \mu_i}\psi_{ij}(U_{ij}, \delta_{ij}; \mu_i)\right]$ and $B(\mu_i) = E[\psi_{ij}(U_{ij}, \delta_{ij}, \mu_i)\psi_{ij}(U_{ij}, \delta_{ij}, \mu_i)^T]$. In our case, $A(\mu_i) = E(W_{ij}) = 1$, and

$$B(\mu_i) = E\left[\frac{\delta_{ij}(Y_{ij} - \mu_i)^2}{K_i^2(Y_{ij})}\right]$$

$$= E\left(E\left[\frac{\delta_{ij}(Y_{ij} - \mu_i)^2}{K_i^2(Y_{ij})} | Y_{ij}\right]\right)$$

$$= E\left(\frac{(Y_{ij}-\mu_i)^2}{K_i^2(Y_{ij})}E(\delta_{ij}|Y_{ij})\right)$$
$$= E\left(\frac{(Y_{ij}-\mu_i)^2}{K_i(Y_{ij})}\right).$$

Therefore as $n_i \to \infty$, $\hat{\mu}_i$ is $\text{AN}\big(\mu_i, \sigma^2_{\hat{\mu}_i}\big)$ where $\sigma^2_{\hat{\mu}_i} = B(\mu_i)/n_i$ and "AN" stands for asymptotic normal.

5.1 WEIGHTED SUM OF SQUARES

In one-way ANOVA, the total corrected sum of squares is decomposed into a deviation of the observations about their group means plus a deviation of the group means about the overall mean (See Section 3.0). For the case of censored log survival data, we define the weighted total corrected sum of squares as follows:

$$TSS_w = \sum_i \sum_j W_{ij}(U_{ij} - \hat{\mu})^2,$$

where $\hat{\mu} = \frac{\sum_i \sum_j W_{ij}U_{ij}}{\sum_i \sum_j W_{ij}}$. Similarly, the usual treatment sum of squares is modified to form its weighted counterpart $SSTR_w = \sum_i \sum_j W_{ij}(\hat{\mu}_i - \hat{\mu})^2$ and the weighted error sum of squares is defined as $SSE_w = \sum_i \sum_j W_{ij}(U_{ij} - \hat{\mu}_i)^2$. Note that the usual identity of decomposition of total sum of squares into treatment and error still follows. That is,

$$\sum_i^p \sum_j^{n_i} W_{ij}(U_{ij} - \hat{\mu})^2 = \sum_i^p \sum_j^{n_i} W_{ij}(U_{ij} - \hat{\mu}_i)^2 + \sum_i^p \sum_j^{n_i} W_{ij}(\hat{\mu}_i - \hat{\mu})^2.$$

This follows since the product term equals zero as shown below:

$$\begin{aligned}\sum_i \sum_j W_{ij}(U_{ij} - \hat{\mu}_i)(\hat{\mu}_i - \hat{\mu}) &= \sum_i (\hat{\mu}_i - \hat{\mu}) \sum_j W_{ij}(U_{ij} - \hat{\mu}_i) \\ &= \sum_i (\hat{\mu}_i - \hat{\mu}) \sum_j (W_{ij}U_{ij} - W_{ij}\hat{\mu}_i) \\ &= \sum_i (\hat{\mu}_i - \hat{\mu}) \left[\sum_j W_{ij}U_{ij} - \sum_j W_{ij}\hat{\mu}_i\right] \\ &= 0\end{aligned}$$

because $\hat{\mu}_i = \sum_j W_{ij}Y_{ij} / \sum_j W_{ij}$.

An alternative expression of $SSTR_w$ is

$$\begin{aligned}SSTR_w &= \sum_i^p \sum_j^{n_i} W_{ij}(\hat{\mu}_i - \hat{\mu})^2 = \sum_i^p (\hat{\mu}_i - \hat{\mu})^2 \sum_j^{n_i} W_{ij} \\ &= \sum_i W_{i.}(\hat{\mu}_i - \hat{\mu})^2 \\ &= \sum_i W_{i.}\hat{\mu}_i{}^2 - \frac{(\sum_i W_{i.}\hat{\mu}_i)^2}{\sum_i W_{i.}},\end{aligned}$$

which has the same expression as in the classical ANOVA except that the role of n_i is replaced by the group-specific sum of weights $W_{i.}$. When the weights within groups are normalized to the group sizes, which naturally occurs when the last observation in the group is a failure and $K_i(.)$ is estimated using (5.3), then the weighted treatment sum of squares can be written as

$$SSTR_w = \sum_i n_i\hat{\mu}_i^2 - \frac{(\sum_i n_i\hat{\mu}_i)^2}{N}. \tag{5.6}$$

Because, similar to the standard analysis of variance, there are p deviations and one degree of freedom is lost because of the constraint that the sum of the deviations

equal 0, the degrees of freedom for the IPCW ANOVA treatment sum of squares remains to be $(p-1)$. Similarly, the IPCW ANOVA error sum of squares is defined as

$$\begin{aligned} SSE_w &= \sum_{i}^{p}\sum_{j}^{n_i} W_{ij}(U_{ij}-\hat{\mu}_i)^2 \\ &= \sum_{i}^{p}\sum_{i}^{n_i} W_{ij}{U_{ij}}^2 - \sum_{i}^{p}\sum_{j}^{n_i}\hat{\mu}_i^2 W_{ij} \\ &= \sum_{i}^{p}\sum_{i}^{n_i} W_{ij}{U_{ij}}^2 - \sum_{i}^{p} W_{i.}{\hat{\mu}_i}^2 \end{aligned}$$

To obtain the error degrees of freedom, we note that associated with the i^{th} component of the error sum of squares, there are $(\sum_j^{n_i}\delta_{ij}-1)$ independent degrees of freedom. Therefore we define the error degrees of freedom (EDF) for the SSE_w as

$$EDF = \sum_i \left(\sum_j \delta_{ij} - 1\right) = n^* - p,$$

where $n^* = \sum_i \sum_j \delta_{ij}$.

5.2 IPCW F RATIO

We propose the following F-ratio to test the hypothesis $Ho : \alpha_1 = ... = \alpha_p$

$$F^w = \frac{SSTR_w/(p-1)}{SSE/(n^*-p)} = \frac{MSTR_w}{MSE_w}.$$

This IPCW F-ratio serves the same purpose as the F-ratio for uncensored data. If all treatment effects $(\alpha_i's)$ are equal, one would expect that the sample group means will be close to each other, and hence the $MSTR_w$ will be small compared to MSE_w.

Similar to the F-ratio in classical ANOVA one would expect that since $MSTR_w$ and MSE_w are approximating means squares due to treatment and error respectively, F^w might follow an F distribution with $(p-1)$ and (n^*-p) degrees of freedom. We investigate this in the simulation study.

5.3 ESTIMATION OF VARIANCE COMPONENTS

For the IPC weighted ANOVA, there are two variance components to be estimated namely σ^2 and $\sigma^2_{\hat{\mu}_i}$. We begin with the estimation of $\sigma^2_{\hat{\mu}_i}$ using the empirical estimate of the quantity $E\left(\frac{(Y_{ij}-\mu_i)^2}{K_i(Y_{ij})}\right)$. More specifically, $\sigma^2_{\hat{\mu}_i}$ can be estimated with the empirical variance estimator as $\sum_i \sum_j \psi^2/N^2$, where $\psi_{ij}(U_{ij}, \delta_{ij}; \mu_i) = W_{ij}(U_{ij}-\mu_i)$, or using the model based estimator as $\sum_i \sum_j \frac{(Y_{ij}-\hat{\mu}_i)^2}{K_i(\hat{Y}_{ij})}/N$.

Then having obtained $\hat{\sigma}^2_{\mu_i}$, the estimation of σ^2 proceeds as in the usual ANOVA fashion by equating observed and expected values of the sum of squares and solving for the estimator. The expectation of the treatment sums of squares and the error sums of squares is as follows

$$
\begin{aligned}
E(SSTR_w) &= E\left[\sum_i W_{i.}\hat{\mu}_i{}^2 - \frac{\left(\sum_i W_{i.}\hat{\mu}_i\right)^2}{\sum_i W_{i.}}\right] \\
&= E\left[\sum_i n_i\hat{\mu}_i{}^2\right] - E\left[\frac{\left(\sum_i n_i\hat{\mu}_i\right)^2}{N}\right] \quad \textit{by (5.6)} \\
&= \sum_i n_i E(\hat{\mu}_i)^2 + \sum_i n_i Var(\hat{\mu}_i) - \frac{1}{N}\left[Var\left(\sum_i n_i\hat{\mu}_i\right) + \left(E\left(\sum_i n_i\hat{\mu}_i\right)^2\right)\right] \\
&= \sum_i n_i\mu_i^2 + \sum_i n_i Var(\hat{\mu}_i) - \frac{1}{N}\left[Var\left(\sum_i n_i\hat{\mu}_i\right) + \left(\left(\sum_i n_i\mu_i\right)^2\right)\right]
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_i n_i Var(\hat{\mu}_i) - \frac{Var \sum_i n_i \hat{\mu}_i}{N} + \left[\sum_i n_i \mu_i^2 - \frac{\left(\sum_i n_i \mu_i\right)^2}{N} \right] \\
&= \sum_i B(\mu_i) - \frac{\sum_i n_i B(\mu_i)}{N} + \left[\sum_i n_i \mu_i^2 - \frac{\left(\sum_i n_i \mu_i\right)^2}{N} \right]. \\
E(SSE_w) &= E\left[\sum_i^p \sum_j^{n_i} W_{ij} U_{ij}^2 - \sum_i^p W_{i\cdot} \hat{\mu}_i^2 \right] \\
&= E\left[\sum_j^{n_i} \frac{\delta_{ij} Y_{ij}^2}{K(Y_{ij})} \right] - E\left[\sum_i^p n_i \hat{\mu}_i^2 \right] \\
&= \sum_i^p \sum_j^{n_i} E\left[E\left(\frac{\delta_{ij} Y_{ij}^2}{K(Y_{ij})} | Y_{ij} \right) \right] - \sum_i^p n_i [E(\hat{\mu}_i)]^2 - \sum_i^p n_i Var(\hat{\mu}_i) \\
&= \sum_i^p \sum_j^{n_i} E[Y_{ij}^2] - \sum_i^p n_i \mu_i^2 - \sum_i^p n_i \sigma_{\hat{\mu}_i}^2 \\
&= N\sigma^2 - \sum_i^p n_i \sigma_{\hat{\mu}_i}^2.
\end{aligned}
$$

Therefore an estimate of σ^2 is $\frac{SSE_w + \sum_i^p n_i \hat{\sigma^2}_{\hat{\mu}_i}}{N}$.

Although unbiasedness of estimators is a desirable property when estimating means. Searle, Casella and McCullogh [46] bring into question the merit of unbiasedness when estimating variances. Particularly, unbiasedness is a property that is established over infinitely many realizations of exactly the same experiment. However, repetition of the same data gathering process with a right censoring mechanism yielding the same pattern of unbalancedness as in the case of the IPCW ANOVA is unrealistic. Even when unbiasedness is achieved, such estimators can yield negative estimators of positive variance components which is not ideal, to say the least.

5.4 CONTRASTS

Using this approximate F-distribution, inference can be made on the IPCW ANOVA model as in the classical model. A rejection of the null hypothesis implies that there is a difference between group means but it does not point to where the difference lies. It is often of interest to identify the pair or group of means that differ. A contrast allows the researcher to test a specific hypothesis of interest rather than the global null hypothesis. Contrast comparisons in the IPCW ANOVA is similar to the standard ANOVA model except that the IPC weighted group means are used instead. The contrast denoted by Ψ defines a specific comparison over group means with the constraint that the weights sum to zero i.e., $\Psi = \sum_i c_i \mu_i$ where $\sum_i c_i = 0$. An approximate unbiased estimator of the contrast Ψ is given by $\hat{\Psi} = \sum_i c_i \hat{\mu}_i$ where,

$$E[\hat{\Psi}] = E\left[\sum_i c_i \hat{\mu}_i\right] = \sum_i c_i E[\hat{\mu}_i] \approx \sum_i c_i \mu_i$$

$$Var[\hat{\Psi}] = Var\left[\sum_i c_i \hat{\mu}_i\right] = \sum_i c_i^2 Var(\hat{\mu}_i).$$

Since $\hat{\Psi}$ is a linear combination of sample group means which are asymptotically normally distributed, $\hat{\Psi}$ itself is also asymptotically normally distributed. Thus any inference for linear combinations of the means in the inverse probability of censoring weighted ANOVA can be conducted similar to that in the standard ANOVA model. For example, a 95% confidence interval for Ψ is given by $\hat{\Psi} \pm t_{n^*-p,\alpha/2} \sqrt{\sum_i c_i^2 \widehat{Var}(\hat{\mu}_i)}$.

5.5 SIMULATION STUDY

We conducted simulations to demonstrate the usefulness of the IPCW ANOVA method in decomposing the total variation into different sources and to examine the properties of the IPCW ANOVA estimator. We compared the IPCW ANOVA method to ANOVA using only the complete cases, the accelerated failure time model (AFT), the Kaplan Meier(KM) and the Buckley James (BJ) method. For the KM method, the restricted mean in each group was calculated as the area under the survival curve for that group. We used the maximum time for all curves as a common upper limit for calculating the area under the survival curve so that the values for the different would be comparable. Data were simulated for subjects in 3 groups of equal sizes. Log failure times Y_{ij} were assumed to follow the model in (3.2) where the $\epsilon'_{ij}s$ are independently and identically distributed normal random variates with mean 0 and variance σ^2. Random censoring times C_{ij}, independent of failure times, were generated from a Uniform(0,τ) distribution, where the value of τ was fixed in advance to control the censoring rates at 25, 35 and 45 percent. The observed data consists of (U_{ij}, δ_{ij}) along with a group membership indicator, where $U_{ij} = \min(T_{ij}, C_{ij})$ and $\delta_{ij} = 1 if T_{ij} < C_{ij}$. Five thousand Monte-Carlo datasets were generated each of size $3n$, where n was varied between 50 and 100 to examine the performance of the estimator at small and moderate sample sizes. For each dataset, we estimated the probability of censoring at each time point using the Kaplan-Meier estimator of the censoring time by considering the censored observations as 'events' and events as 'censored'. Inverse probability of censoring weights for each observation was calculated as the event status indicator divided by the Kaplan-Meier estimate. Thus observations with true event times received a weight equal to the inverse of the KM estimator of censoring times and censored observations received a weight of 0.

We then compared the type I error rates under the null hypothesis of equal means of 4 of the IPCW ANOVA model to the ANOVA model using the complete cases alone, the lognormal AFT model, the Buckley James(BJ) procedure and the Kaplan Meier(KM) method. We set $\sigma^2 to\ be 1, 1.5 and 2$. Type I error rates for the IPCW ANOVA was calculated by comparing the proposed empirical F-ratio to its proposed reference F-distribution. For the ANOVA model using only the complete cases, the type I error rate was obtained by comparing the F-ratio to its reference F-distribution. Type I error rates for the AFT were obtained using a likelihood ratio test statistic compared to its reference χ^2 distribution. Type I error rates for the BJ and KM methods were calculated using the Wald chi-square test based on the group means and standard error of the estimates obtained from these procedures. We examined the accuracy of the estimated group means of the IPCW ANOVA model, AFT model and the KM method in comparison to the truth under the null hypothesis of equal group means of 5 and under the alternative hypothesis of group means of 3, 5, and 8 respectively. We set the sample size in each group, n_i, at 50 and $\sigma^2 = 2.25$.

The power of the IPCW ANOVA model to test the null hypothesis of equal group means against the pre-specified alternative hypothesis of group means equal to 4, 3, and 4 was compared to that of the AFT model.

We also examined the performance of the IPCW ANOVA model under a mis-specified model. The model is mis-specified because the covariates are not linearly related to the mean of the response and because the underlying distribution of the log of the survival times is not normal. Here data were generated according to a proportional hazards framework. More specifically, the hazard function of the j^{th} person in the i^{th} group follows the hazard function $h_{ij}(t) = h_0(t)exp(\beta_i)$, where h_0 is the baseline hazard function at time t. We assumed a baseline hazard of 0.5 and set

the sample size in each group, n_i, in each group at 50. Event times were generated from an exponential distribution with rate equal to the hazard. The mean of the exponential distribution is given by $1/h_{ij}(t)$. First we generated equal group means by setting ($\beta_1 = \beta_2 = \beta_3 = 0$) so that the mean in each group was 1/0.5=2. We also simulated different means in each group by setting ($\beta_1 = 0.5, \beta_2 = -0.25, \beta_3 = 0.4$) so that the mean in each of the three groups 1,2,3 were ($1/0.5 = 2, 1/(0.5e^{0.25}) = 2.57$, and $1/(0.5e^{0.4}) = 1.34$) respectively. Under the mis-specified model, censoring times were generated from a $Uniform(0, \tau)$, where the value of τ was used to control the censoring rates at 25%. The observed data consists of (U_{ij}, δ_{ij}) along with a group membership indicator, where U_{ij} and δ_{ij} are as defined previously.

5.6 SIMULATION RESULTS

Table (5.1) shows the type I error rates of the IPCW ANOVA, CC ANOVA, lognormal AFT, BJ, and KM models. The test based on the IPCW F-ratio maintained the type I error rate, while the error rate based on the CC ANOVA model and the AFT model were slightly above the nominal rate of 0.05. In general, the IPCW ANOVA was more conservative than either the other models irrespective of sample size and censoring percentage. The BJ procedure did not always converge. In some cases, no convergence was obtained but an average value of the estimated means was returned and at other times no convergence was obtained and the cycle failed completely returning no value for the estimated means. The number of iterations for which we were to calculate the BJ estimates ranged from 287 to 5000. Therefore a comparison between the type I error rate of the BJ type method and other methods can not be made. For this reason, the BJ method was not used in further comparisons.

Table 5.1: Type I error rate for testing $H_0 : \mu_1 = \mu_2 = \mu_3 = 4$ under different scenarios. The data were generated from log-normal distributions

		n=50			n=100		
	Censoring rate	25%	35%	45%	25%	35%	45%
	CC	0.054	0.053	0.053	0.049	0.049	0.047
	IPCW ANOVA	0.043	0.039	0.040	0.038	0.036	0.037
σ=1	AFT	0.055	0.059	0.057	0.054	0.052	0.052
	BJ	0.048	0.045	0.039*	0.043	0.037	0.037
	KM	0.064	0.065	0.05	0.058	0.057	0.055
	CC	0.055	0.055	0.051	0.047	0.044	0.045
	IPCW ANOVA	0.039	0.043	0.043	0.035	0.035	0.037
σ=1.5	AFT	0.057	0.053	0.058	0.054	0.052	0.05
	BJ	0.041	0.039*	0.031*	0.042	0.037*	0.033*
	KM	0.064	0.063	0.067	0.058	0.055	0.054
	CC	0.054	0.051	0.059	0.047	0.043	0.047
	IPCW ANOVA	0.036	0.038	0.054	0.035	0.033	0.048
σ=2.0	AFT	0.060	0.055	0.056	0.055	0.05	0.049
	BJ	0.042	0.042*	0.053*	0.041	0.038*	0.035*
	KM	0.066	0.061	0.064	0.051	0.059	0.055

*=not based on 5000 iterations

CC=Complete Case,AFT=Accelerated failure time model, BJ=Buckley James, KM=Kaplan-Meier, IPCW=IPCW ANOVA

We also dropped the complete case analysis from further comparisons because a complete case analysis is inefficient and does not maximize the use of available data since subjects with missing data are deleted and their data information is not utilized in the analysis.

The estimated group means, model variance and Monte Carlo variance, which is an approximation of the variance by the sample variance of the estimates obtained from all the stochastic simulations,are given in Table (5.2). The group means were consistently estimated by the IPCW ANOVA model. This was true both at low(25%) and moderate(45%) levels of censoring. The bias in the estimated means using the KM method was slightly worse than either the IPCW or AFT model when group means were unequal. The estimated Monte Carlo variance of the IPCW ANOVA model was comparable to those of the AFT and KM models when the censoring rate was low (25%) but was slightly greater than either the AFT or the KM at moderate(45%) levels of censoring, except at 45% censoring in group 3. The estimated model variance of the IPCW estimator was comparable to its Monte Carlo counterparts.

Under model mis-specification, all three methods had a considerable amount of bias in the estimated means. The IPCW ANOVA model under estimated the group means compared to the other methods. This is understandable since we did not take into consideration when estimating the weights the proportionality of hazards between groups. The Monte Carlo variance of the IPCW estimator was higher than either the AFT or the KM. The model based variance of the IPCW estimator was much lower than its MC counterparts.

The power of the IPCW F-ratio compared to all other models to detect the specified alternative is shown in Figure (5.1).

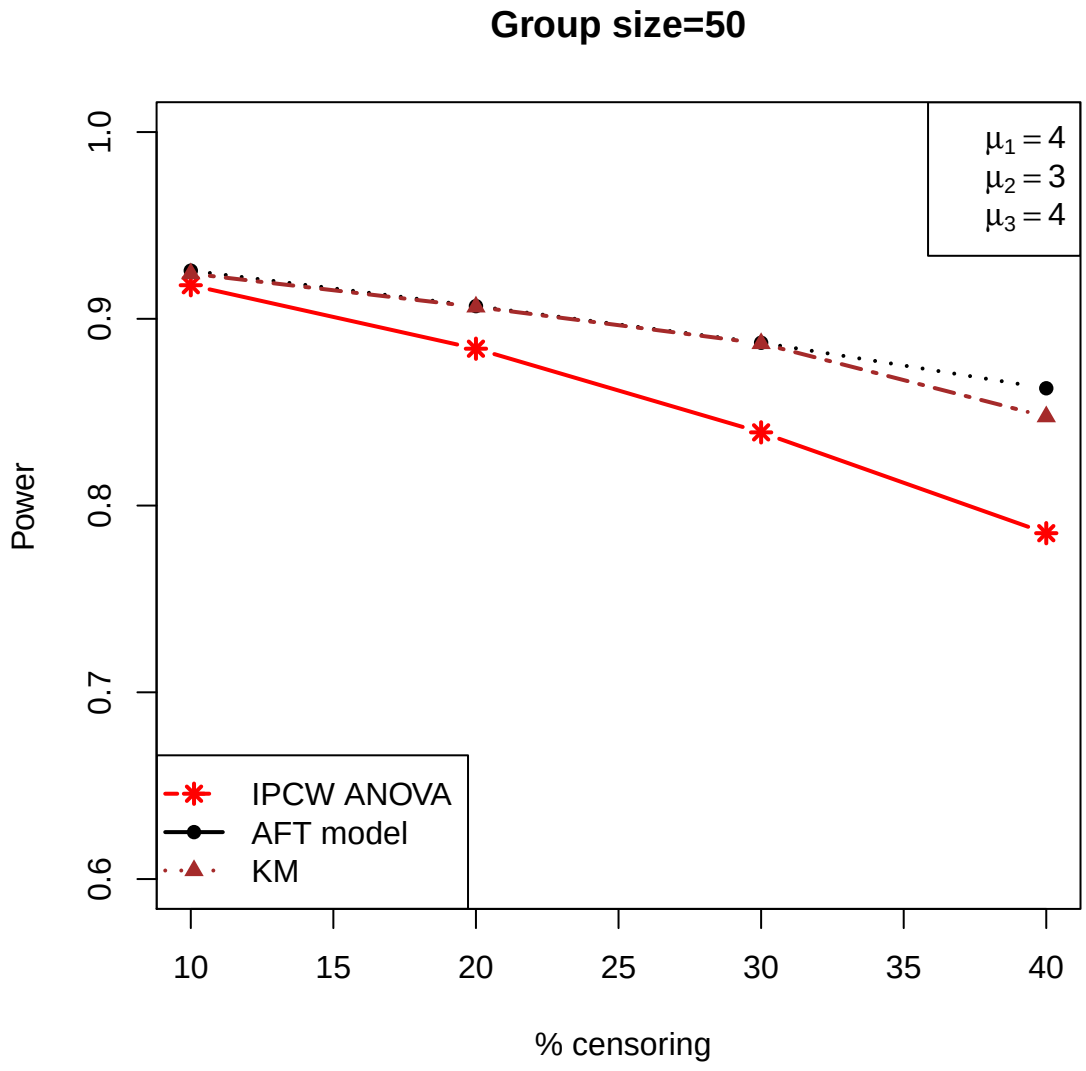

Figure 5.1: Empirical power curves for testing $H_0 : \mu_1 = \mu_2 = \mu_3$ against pre-specified alternative $\mu_1 = 4$, $\mu_2 = 3$, $\mu_3 = 4$. Failure times are generated from the log-normal distribution.

Table 5.2: Consistency and efficiency of group mean estimates in IPCW ANOVA. $n_i = 50, \sigma^2 = 2.25$.

		Group 1			Group 2			Group 3		
% censoring	Method	Estimate	MC Var	Var	Estimate	MC Var	Var	Estimate	MC Var	Var
		$\hat{\mu}_1$		$\hat{\sigma}^2_{\hat{\mu}_1}$	$\hat{\mu}_2$		$\hat{\sigma}^2_{\hat{\mu}_2}$	$\hat{\mu}_3$		$\hat{\sigma}^2_{\hat{\mu}_3}$
25% H_0	TRUE	5			5			5		
	AFT	5.00	0.057		5.00	0.057		5.00	0.058	
	KM	4.99	0.056		4.99	0.056		4.99	0.056	
	IPCW	4.99	0.056	0.052	4.99	0.056	0.052	4.99	0.056	0.052
25% H_1	TRUE	3			5			8		
	AFT	3.00	0.051		5.00	0.057		8.00	0.070	
	KM	3.01	0.051		5.00	0.058		7.38	0.126	
	IPCW	3.00	0.052	0.051	4.99	0.059	0.058	7.99	0.073	0.072
45% H_0	TRUE	5			5			5		
	AFT	5.00	0.071		5.00	0.070		5.00	0.070	
	KM	4.98	0.070		4.97	0.067		4.98	0.068	
	IPCW	4.97	0.079	0.080	4.96	0.077	0.080	4.964	0.077	0.079
45% H_1	TRUE	3			5			8		
	AFT	3.00	0.052		5.00	0.066		8.01	0.114	
	KM	3.01	0.057		4.99	0.072		7.16	0.172	
	IPCW	2.99	0.060	0.060	4.96	0.080	0.080	7.77	0.150	0.136

AFT=Accelerated failure time model, KM=Kaplan-Meier, IPCW=IPCW ANOVA

MC=Monte Carlo Variance, Model= Variance estimate

$\sigma^2_{\hat{\mu}_i}$ was estimated using $\frac{1}{n_i}\sum \frac{(Y_{ij}-\hat{\mu}_i)^2}{\hat{K}_i(Y_{ij})}$

Table 5.3: Consistency and efficiency of group mean estimates under model misspecification.

		Group 1			Group 2			Group 3		
% censoring	Method	Estimate	MC Var	Var	Estimate	MC Var	Var	Estimate	MC Var	Var
		$\hat{\mu}_1$		$\hat{\sigma}^2_{\hat{\mu}_1}$	$\hat{\mu}_2$		$\hat{\sigma}^2_{\hat{\mu}_2}$	$\hat{\mu}_3$		$\hat{\sigma}^2_{\hat{\mu}_3}$
25% H_0	TRUE	2			2			2		
	KM	1.86	0.061		1.86	0.062		1.85	0.060	
	AFT	1.91	0.075		1.91	0.077		1.90	0.074	
	IPCW	1.74	0.089	0.078	1.74	0.089	0.077	1.73	0.084	0.076
25% H_1	TRUE	2			2.57			1.34		
	KM	1.91	0.079		2.29	0.093		1.35	0.052	
	AFT	1.86	0.061		2.31	0.097		1.35	0.033	
	IPCW	1.74	0.089	0.044	2.23	0.147	0.073	1.12	0.038	0.020

AFT=Accelerated failure time model, KM=Kaplan-Meier, IPCW=IPCW ANOVA

MC=Monte Carlo Variance, Model= Model based variance

The IPCW, Kaplan-Meier and AFT model had greater than 80% power to detect the prespecified alternative. As the censoring rate increases, the power of all three methods declines. The power of the KM method is comparable to that of the AFT model while the IPCW ANOVA model under performs in comparison to both the AFT and KM models. However, it should be noted that the data were generated from a lognormal distribution and a lognormal distribution was specified as the underlying distribution in the AFT model. Therefore, the AFT model might be expected to perform better since the true underlying distribution was correctly specified. Figure (5.6) shows the power curves at different censoring rates for all models when data were generated from an exponential hazard distribution. Here the IPCW outperforms all other models and maintains at least an 80% percent power to detect the prespecified alternative.

5.7 REAL DATA ANALYSIS 1

We applied the methods on a study from the North Central Cancer Treatment Group (NCCTG) in 228 patients with advanced lung cancer. Patients were evaluated on how well they performed usual daily activities. Performance scores along with other variables related to caloric intake and weight loss were used as predictors of survival. In this dissertation, the goal of the analysis is to examine the effect of age greater than or equal to 65 on survival time. We begin by examining the normality assumption in each of the groups. Figure (5.7) show that the log transformation successfully normalized the sample for age < 65.

The survival curves for the two age groups are shown in Figure (5.7). There is a rapid decline in the survival probabilities in both age groups after a log time of

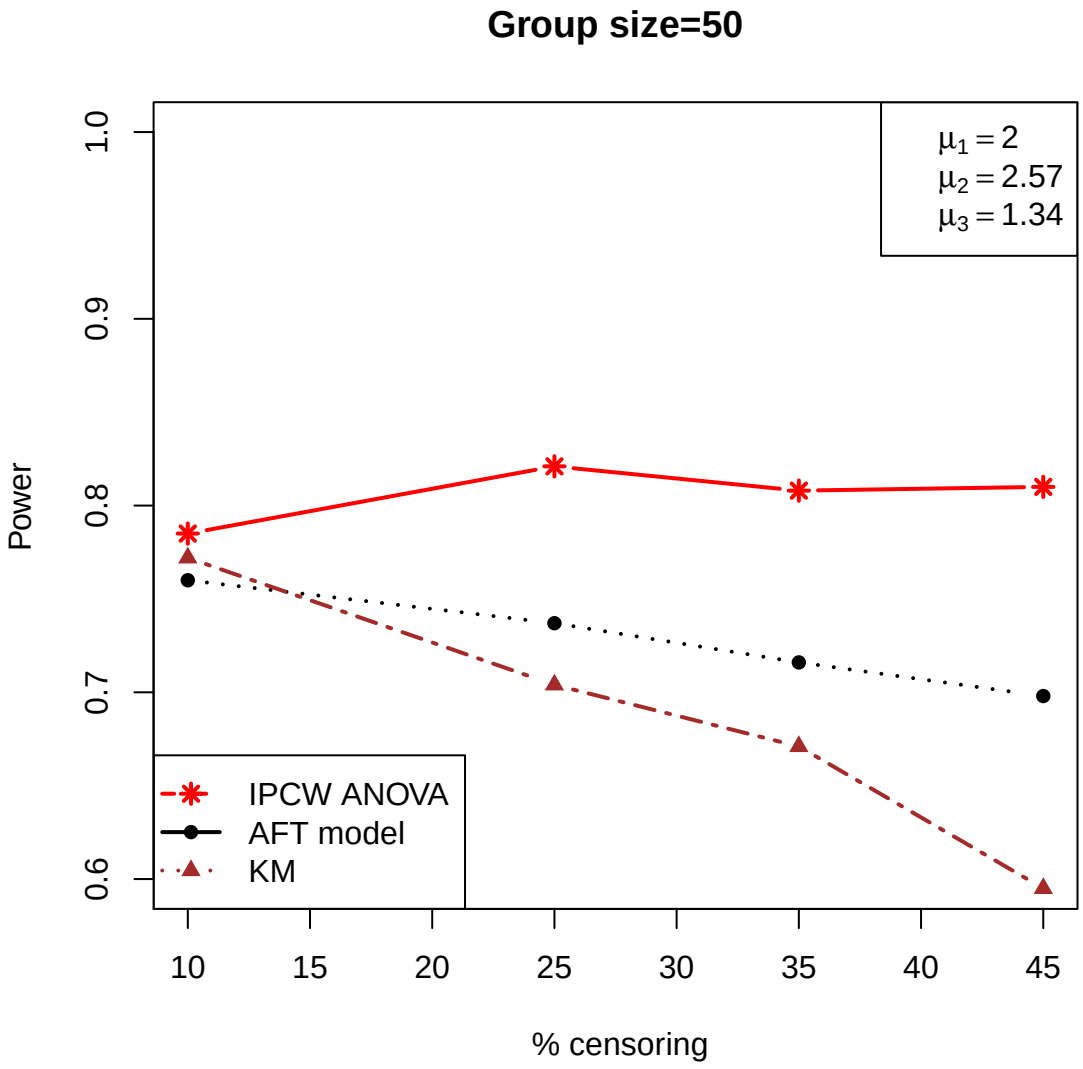

Figure 5.2: Empirical power curves for testing $H_0 : \mu_1 = \mu_2 = \mu_3$ against pre-specified alternative $\mu_1 = 2$, $\mu_2 = 2.57$, $\mu_3 = 1.34$. Failure times are generated from an exponential hazard distribution

5. In general, younger patients with age less than 65 have a better overall survival experience. The mean survival times for patients under 65 and those 65 or older using the IPCW ANOVA IPCW was 6.70(0.08) and 6.30(0.12) respectively.

Table (5.5) shows the IPCW ANOVA decomposition of sums of squares. The model is significant ($F = 7.09, p = 0.009$). Age less than 65 accounts for 4.2% of the variation in the data ($R^2 = 0.042$).

5.8 REAL DATA ANALYSIS 2

We also applied our methods to the publicly available clinical trial data collected by the Radiation Therapy Oncology Group (RTOG) (Kabfleisch & Prentice, 2002). The trial compared the efficacy in improving survival of two treatment options: radiation therapy alone or a combination of radiation therapy and a therapeutic agent in patients with squamous cell carcinoma of the mouth and throat. The objective of the study was to compare the two treatments with respect to patient survival. Survival is measured as the time in days from diagnosis. Patients are classified according to the stage of the primary tumor (T stage). The classification of patients according to the stage of the primary tumor is as follows: 1: $\leq 2cm$ in diameter, 2: $2 - 4cm$ in diameter, 3: $> 4cm$ in diameter and 4: massive invasive tumor. Survival curves of these four groups are presented in Figure 5.8. Approximately 30% of the survival times are censored and there is a great degree of heterogeneity among the patients. Specifically, 50% of the 6 patients in group 1, 62.5% of the 26 patients in group 2, 43% of the 93 in group 3 and 22% of the 97 patients in group 4 are censored. We begin by first checking the assumptions for the ANOVA test. Because the observed survival times were not normally distributed, a log transformation was applied to

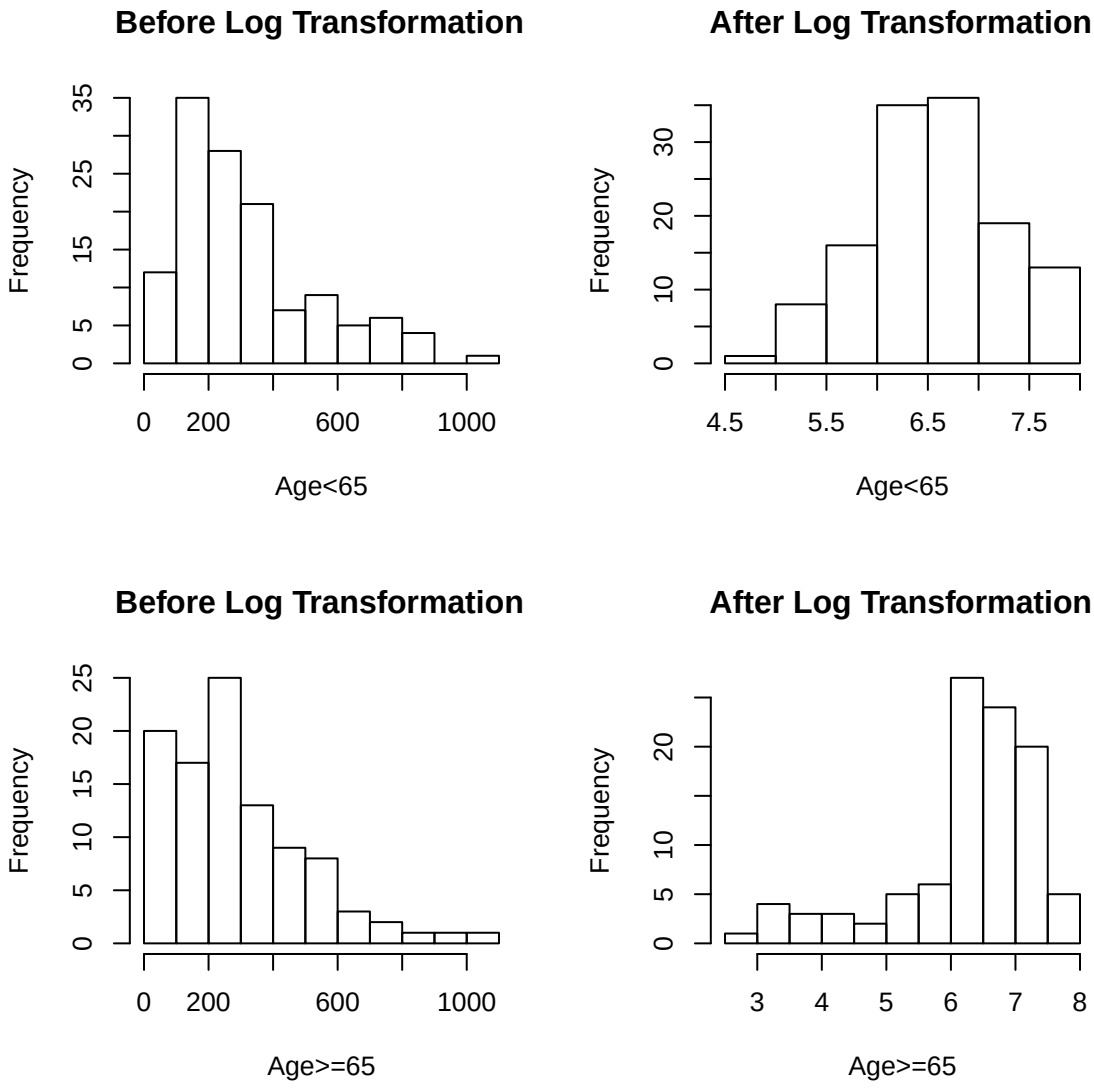

Figure 5.3: Log Transformation of the NCCTG Lung Cancer Data

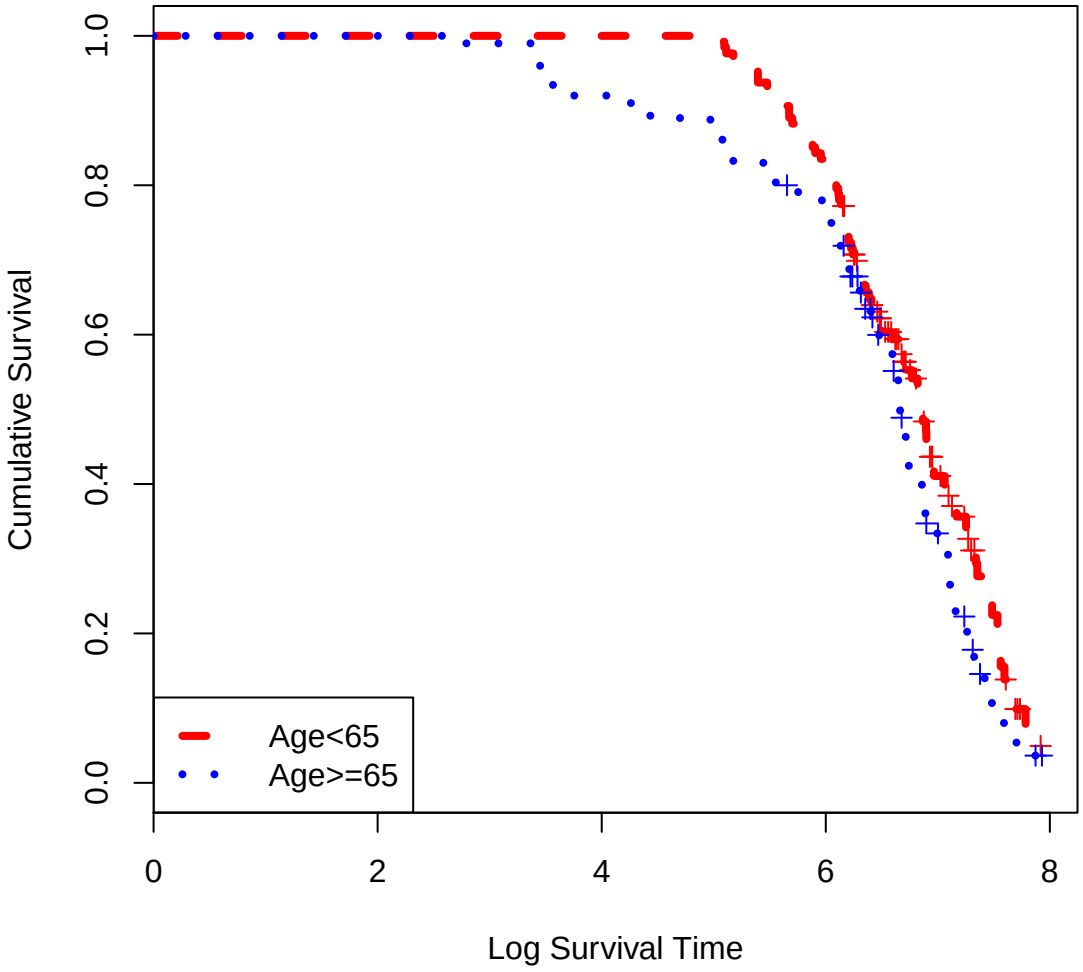

Figure 5.4: Survival Curve of the NCCTG Lung Cancer Data

Table 5.4: Inverse probability of censoring weighted analysis of variance for the NCCTG Lung Cancer data

Source	df	Sum of Squares (SS)	Mean Squares (MS)	F	$Pr > F$
Model	1	8.505	8.505	7.09	0.009
Error	163	195.530	1.200		
Corrected Total	164	204.305			

Rsquare: 0.042

the data. Q-Q plots (not shown) indicate that the data are normally distributed after transformation. It is expected that the stage of the disease would be related to survival experience, thus comparison of mean survival across T-stage classification is of interest. In addition, we are also interested in testing if there are differences in the survivorship experience between patients classified as having massive invasive tumor and those with less invasive tumors (T=1,2,3). The median survival times of the logged survival times are 6.24, 6.66, 6.14 and 5.76 respectively. The results of the test of survivorship experience among the four disease groups and the test of the difference in the survivorship experience between patients with massive invasive tumor vs. less invasive tumor using the IPCW ANOVA model is shown.

The Kaplan-Meier curves in Figure (5.8) show differences in the survivorship experience between the groups. Patients with massive invasive tumor appear to have the worst survival experience. Patients with a primary tumor measuring $2 - 4cm$ in diameter appear to have the best overall survival experience. However, the survival curves of the groups cross at various points indicating no difference in the overall

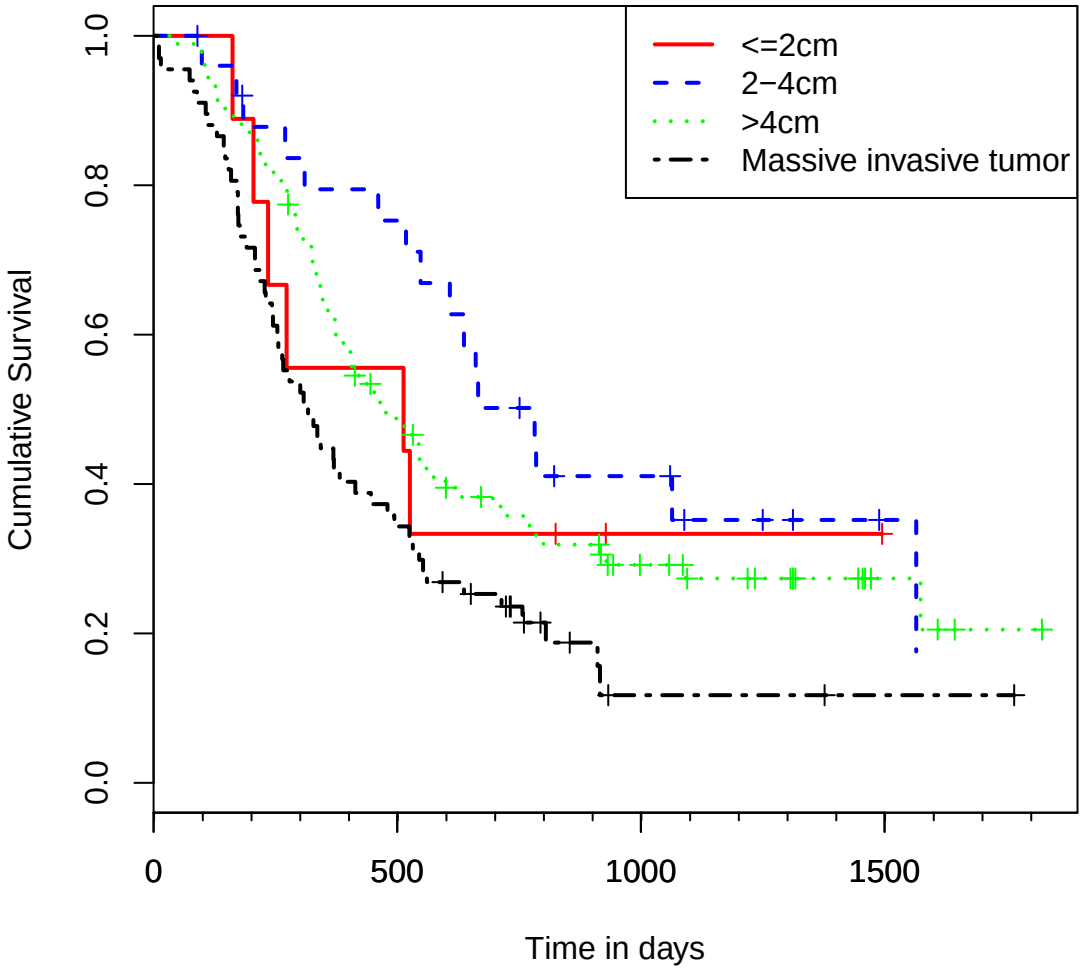

Figure 5.5: Kaplan-Meier Survival Curves for the RTOG data

Table 5.5: Inverse probability of censoring weighted analysis of variance using the RTOG data.

Source	df	Sum of Squares (SS)	Mean Squares (MS)	F	$Pr > F$
Model	3	12.305	4.102	5.087	0.002
Error	138	111.267	0.806		
Corrected Total	141	123.572			

Rsquare: 0.1

survival experience. The model rejects the null hypothesis of no difference in the group means. The p-value of the test was below 0.05. Table 5.5 shows the IPCW ANOVA decomposition of sums of squares. The model is significant ($F = 5.09, p = 0.002$). T stage classification alone accounts for 10% of the variation in the data ($R^2 = 0.1$). A hypothesis of interest might be to examine the survivorship experience between patients with massive invasive tumor vs. less invasive tumors, the contrast results from the IPCW ANOVA yields an estimated difference of about 2.32. This difference was statistically significant in a t-test with 138 degrees of freedom.

6.0 TWO-WAY CROSS CLASSIFIED ANALYSIS OF VARIANCE

Factorial designs are often considered when two or more treatments with two or more levels e.g dose levels are of interest either alone or in combination. Usually, patients are assigned equally to each possible combination of treatment levels resulting in a balanced design. The aim is to study the effect of levels of each treatment by pooling across all other treatments. The use of factorial designs in clinical trials with time to event outcomes has been studied by Akritas and LaValley[1], who proposed a non parametric approach to the analysis of factorial designs with censored data without making any assumptions on the interaction effects. Slud[49] discussed methodological issues that arise in factorial designs in survival experiments with two binary covariates under the proportional hazards framework.

The goal here, as in the one factor case, is to extend the general linear model to the censored data setting when a factorial design is employed. The two way analysis of variance is an extension of the one way ANOVA model to the case where data are classified by two qualitative variables. Let Y_{ijk} be the k^{th} observation in the i^{th} level of factor A and the j^{th} level of factor B. The complete model for the cross classified two way ANOVA model is given by:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk} \quad i = 1 \cdots a, j = 1 \cdots b, k = 1 \cdots n_{ij}. \tag{6.1}$$

where $\epsilon_{ijk} \sim N(0, \sigma^2)$, μ is the overall grand mean, α_i is the fixed effect of the row factor A, β_j is the fixed effect of the column factor B and $(\alpha\beta)_{ij}$ is the interaction of the row and column factors. Typically, interest lies in testing the effect of the interaction term and/or the effect of the row or column factors. Analysis of variance tests can then be formulated in terms of contrasts of parameters or equivalently in terms of full and reduced models.

Assuming the data are balanced i.e n_{ij}'s are all equal, the total sums of squares is still a sum of the sums of squares of the individual sources of variation in the data. That is

$$\begin{aligned}\sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y_{...}})^2 = &\sum_i \sum_j \sum_k (\bar{Y_{i..}} - \bar{Y_{...}})^2 \\ &+ \sum_i \sum_j \sum_k (\bar{Y_{.j.}} - \bar{Y_{...}})^2 \\ &+ \sum_i \sum_j \sum_k (\bar{Y_{ij.}} - \bar{Y_{i..}} - \bar{Y_{.j.}} + \bar{Y_{...}})^2 \\ &+ \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y_{ij.}})^2. \end{aligned} \tag{6.2}$$

The reduced model for testing the interaction term is

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk} = \mu_{ij}^1 + \epsilon_{ijk} \tag{6.3}$$

and the full model is the same as that given in (6.1). Then the sums of squares for testing the interaction term is given by

$$SS^1 = SSE^1 - SSE,$$

where $SSE^1 = \sum_i \sum_j \sum_k (Y_{ijk} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j)^2$ is the sums of squares from the reduced model and $SSE = \sum_i \sum_j \sum_k (Y_{ijk} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j - \widehat{(\alpha\beta)}_{ij})^2$ is the sums of squares

from the full model. Then under the null hypothesis of no interaction effect,

$$MS^1 = \frac{SS^1}{df1} \sim \sigma^2 \frac{\chi^2_{df1}}{df1},$$

where $df1 = (a-1)(b-1)$ is the degrees of freedom for the interaction term. Similarly, the mean square error from the full model is $MSE \sim \sigma^2 \frac{\chi^2_{dfE}}{dfE}$. Therefore,

$$F^1 = \frac{MS^1}{MSE} \sim F_{df1,dfE}$$

where $dfE = N - ab$ and $N = \sum n_{ij}$.

In the absence of an interaction effect, a similar test for main effects can be obtained. In this case, the reduced model for testing the main effect of the row factor A is given by

$$Y_{ijk} = \mu + \beta_j + \epsilon_{ijk} = \mu^2_{ij} + \epsilon_{ijk}$$

and the full model is the same as that given in (6.3). The sums of squares for testing the effect of the row factor is given by

$$SS^2 = SSE^2 - SSE^1.$$

With similar arguments to the interaction test, under the null hypothesis of no effect of the row factor,

$$F^2 = \frac{MS^2}{MS^1} \sim F_{df2,df1},$$

where $MS^2 = \frac{SS^2}{df2}$ and $SS^2 = \sum_i \sum_j \sum_k (Y_{ijk} - \hat{\mu} - \hat{\beta}_j)^2$ and $df2 = (a-1)$.
With balanced data, the analysis of variance table is then given by Table (6.1) but when the cross classified data are unbalanced, ANOVA tables can be constructed in many different ways, however, the factorization of total sum of squares into components is not as straightforward as in Equation (6.2).

Table 6.1: Two-way ANOVA Table

Source	df	Sum of Squares (SS)	Mean Squares (MS)	F
Factor A	(a-1)	SSA=$bn\sum_i^a(\bar{Y}_{i..}-\bar{Y}_{...})^2$	MSA=$\frac{SSA}{a-1}$	$F_A=\frac{MSA}{MSE}$
Factor B	(b-1)	SSB=$an\sum_i^b(\bar{Y_{j..}}=\bar{Y}_{...})^2$	MSB=$\frac{SSB}{b-1}$	$F_B=\frac{MSB}{MSE}$
Interaction	(a-1)(b-1)	SSAB=$n\sum_i\sum_j(\bar{Y_{ij.}}-\bar{Y_{i..}}-\bar{Y_{.j.}}-\bar{Y_{...}})^2$	MSAB=$\frac{SSAB}{(a-1)(b-1)}$	$F_{AB}=\frac{MSAB}{MSE}$
Error	N-ab	SSE=$\sum_i\sum_j\sum_k(Y_{ijk}-\bar{Y_{ij.}})^2$	MSE=$\frac{SSE}{N-ab}$	

For survival experiments with time to event outcomes, unbalanced data may be the rule rather than the exception because even when the study is balanced by design, random censoring may introduce imbalance into the data. When data are unbalanced, the regression sums of squares no longer partitions into independent pieces corresponding to main effects and interactions. Furthermore certain hypothesis such as differences in row/column effects $\alpha_i - \alpha_k$ or equality of row factors $H : \alpha_i$ *all equal* are not estimable because the model is over parameterized i.e there are more parameters than can be uniquely estimated from the data. Therefore, similar to the one factor case, we use instead the factorial cell means model since it yields a full rank model so that model parameters are uniquely estimated. The cell means model for the two way cross classified model is given by

$$Y_{ijk} = \mu_{ij} + \epsilon_{ijk} \quad i = 1 \cdots a, j = 1 \cdots b, k = 1 \cdots n_{ij}. \tag{6.4}$$

Here μ_{ij} is the population cell mean of the i, j^{th} combination of factors A and B. In this case, the usual hypothesis of interest may be carried out using the partial F test. Although an ANOVA table may be defined using partial (Type III) sums of squares which represent the improvement in fit when an effect is added last to the model containing all other effects, for this unbalanced two way cross classified model, we will focus on comparing the mean survival time and describing the variation in the data using appropriate sums of squares.

6.1 THE TWO WAY IPCW ANOVA MODEL

The model under consideration here is the two way cross classified cell means model given by

$$Y_{ijk} = log(T_{ijk}) \quad = \quad \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk} = \mu_{ij} + \epsilon_{ijk} \tag{6.5}$$

where T_{ijk} is the survival time of the k^{th} subject in the $(i,j)^{th}$ cell, C_{ijk} is the censoring time of the k^{th} subject in the $(i,j)^{th}$ cell and $\epsilon_{ijk} \sim N(0, \sigma^2)$. The observed data consists of $U_{ijk} = min(T_{ijk}, C_{ijk})$ and δ_{ijk} which is equal to 1 if $T_{ijk} < C_{ijk}$, 0 otherwise.

Following the one way IPCW ANOVA , let $K_{ij}(Y_{ijk}) = \Pr(lnC_{ijk} > lnT_{ijk})$, n_{ij} be the number of observations in cell (i,j) and $N = \sum_i \sum_j n_{ij}$. The weighted objective function to be minimized is:

$$\begin{aligned} Q^* &= \sum_i^a \sum_j^b \sum_k^{n_{ij}} \frac{\delta_{ijk}}{K_{ij}(U_{ijk})} (U_{ijk} - \mu_{ij})^2 \\ &= \sum_i^a \sum_j^b \sum_k^{n_{ij}} W_{ijk} (U_{ijk} - \mu_{ij})^2. \end{aligned}$$

where $W_{ijk} = \frac{\delta_{ijk}}{K_{ij}(U_{ijk})}$. Then for any cell i, j,

$$\hat{\mu}_{ij} = \frac{\sum_k W_{ijk} U_{ijk}}{\sum_k W_{ijk}} = \frac{\sum_k W_{ijk} U_{ijk}}{W_{ij.}} = \frac{\sum_k W_{ijk} U_{ijk}/n_{ij}}{W_{ij.}/n_{ij}} = (\bar{W}_{ij})^{-1} \frac{\sum_k W_{ijk} U_{ijk}}{n_{ij}},$$

where $W_{ij.}$ is the sum of the weights of the $(i,j)^{th}$ cell. It is easy to see that $\hat{\mu}_{ij}$ is the solution to

$$\sum_k^{n_{ij}} \Psi_k(U_{ijk}, \delta_{ijk}, \mu_{ij}) = W_{ijk}(U_{ijk} - \mu_{ij}) = 0.$$

Then using the results established in Lemma 1 of page 20, the asymptotic properties of the weighted estimator can be derived by showing that in this case,

$$\begin{aligned}
A(\mu_{ij}) &= -E\left[\frac{\partial}{\partial\mu_{ij}}\Psi_k(Y_{ijk},\delta_{ijk},\mu_{ij})\right] \\
&= -E\left[\frac{\partial}{\partial\mu_{ij}}W_{ijk}(Y_{ijk}-\mu_{ij})\right] \\
&= E(W_{ijk}) \\
&= 1. \\
B(\mu_{ij}) &= E\left[\Psi_k(Y_{ijk},\delta_{ijk},\mu_{ij})(\Psi_k(Y_{ijk},\delta_{ijk},\mu_{ij}))^T\right] \\
&= E\left[\frac{\delta_{ijk}(Y_{ijk}-\mu_{ij})^2}{K_{ij}^2(Y_{ijk})}\right] \\
&= E\left[\frac{(Y_{ijk}-\mu_{ij})^2}{K_{ij}^2(Y_{ijk})}E\left(\delta_{ijk}|Y_{ijk}\right)\right] \\
&= E\left[\frac{(Y_{ijk}-\mu_{ij})^2}{K_{ij}(Y_{ijk})}\right] \\
&= \hat{\sigma}^2_{\hat{\mu}_{ij}}
\end{aligned}$$

Therefore $\hat{\mu}_{ij} \sim \text{AN}\left(\mu_{ij}, \sigma^2_{\hat{\mu}_{ij}}/n_{ij}\right)$. An empirical estimator of $\sigma^2_{\hat{\mu}_{ij}}$ is $\frac{1}{N}\sum_i\sum_j\sum_k \frac{(Y_{ijk}-\mu_{ij})^2}{K_{ij}(Y_{ijk})}$.

6.2 TESTS OF HYPOTHESIS

Usual hypothesis of interest in factorial designs include a test for the presence of an interaction effect and the effects of the main factors by which the data are crossed in the absence of an interaction. Using regression models and accounting for censored observations using weighting techniques, these tests can be accomplished by comparing the between factor sums of squares to the within(error) sums of squares

for the full and reduced models. Details of the IPCW ANOVA factorial method for testing specific hypothesis of interest are given in the sections that follow.
The hypothesis of no interaction is given by

$$H^1 : \mu_{ij} - \mu_{i^*j} = \mu_{ij^*} - \mu_{i^*j^*} for\ all\ i \neq i^*\ and\ j \neq j^*.$$

The reduced model for testing this hypothesis is

$$Y_{ijk} = \mu_{ij}^1 + \epsilon_{ijk}, \tag{6.6}$$

where $\mu_{ij}^1 = \mu + \alpha_i + \beta_j$, which is the same as the full model given in Equation (6.5) minus the interaction term. The weighted sums of squares for the reduced and full models are given by $SSE^1 = \sum_i \sum_j \sum_k W_{ijk}(U_{ijk} - \hat{\mu}_{ij}^1)^2$ and $SSE = \sum_i \sum_j \sum_k W_{ijk}(U_{ijk} - \hat{\mu}_{ij})^2$ respectively, where the weights W_{ijk} are calculated separately as the survival function of the censoring times in each cell. Then the sums of squares for testing the interaction term is $SS^1 = SSE^1 - SSE$. We propose the following F-statistic for testing H^1:

$$F_{ipcw}^1 = \frac{SS^1/df1}{SSE/dfE} = \frac{MS^1}{MSE}$$

where $df1 = (a-1)(b-1)$ and $dfE = N^* - ab$ are the degrees of freedom associated with the model and error respectively and $N^* = \sum_i \sum_j \delta_{ij}$, which is the total number of uncensored observations in all cells combined.

When there is an interaction present, then the effect of one factor on the response depends on the level of the cross factor and it is unwise to test for main effects since it is unclear how to separate out the effect of a factor from the interaction. In this case, contrast comparisons between cell means is a more useful goal. In the absence of an interaction, the main effects fully explain the data and the full model for testing

any of the main effects is (6.6) The hypothesis of no main effect for any factor, say the row factor A is given by

$$H^2 : W_{i..}\bar{\mu_{i.}} = W_{i^*..}\bar{\mu_{i^*.}} for\ all\ i \neq i^*,$$

where $W_{i..}$ is the total weight in the i^{th} row after summing over all columns. Then the reduced model for testing H^2 is given by

$$Y_{ijk} = \mu^2_{ij} + \epsilon_{ijk} = \mu + \beta_j + \epsilon_{ijk}.$$

Similarly, the sums of squares for the reduced model is given as $SSE^2 = \sum_i \sum_j \sum_k W_{ijk}(U_{ijk} - \mu^2_{ij})^2$. and the proposed F statistic for testing H^2 is

$$F^2_{ipcw} = \frac{SS^2/df2}{SSE^1/dfE},$$

where $df2 = (a-1)$ and dfE is as defined previously.

6.3 SIMULATION STUDY

We conducted simulations to evaluate the performance of the IPCW ANOVA model in a 2×3 factorial design and compared it with the AFT model. Data was generated from a lognormal AFT model using the model $Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$ when an interaction effect was present and $Y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}$ for the main effects model, where ϵ_{ij} was normally distributed with mean 0 and variance 5. The row factor A had 3 levels while the column factor B had 2 levels. We set $\mu = 10$ and the effect of the factors A,B depended on whether the true model was a main effects model or an interaction model. Censoring times were generated from a $Uniform(0, \tau)$, where

Table 6.2: Cell means for specified alternative hypothesis

	Null Model				Main Effects Model				Interaction Model			
	A1	A2	A3	Row Avg	A1	A2	A3	Row Avg	A1	A2	A3	Row Avg
B1	10	10	10	10	8.5	11	13.5	10	8	10	19.5	12.5
B2	10	10	10	10	6.5	9	11.5	9	5	6	11.5	7.5
Col Avg	10	10	10	10	7.5	10	12.5	10	6.5	8	15.5	10

the value of τ was used to control the desired censoring rate at 25% and 45%. The number of observations in each cell was varied between 10,20 and 30 observations before censoring. The results are based 1000 replications.

Table 6.2 shows the true cell means in each of the models. We compared the type I error rates and power of the IPCW ANOVA model against a prespecified alternative to that of the lognormal AFT model. Table 6.4 shows the results of these comparisons. The null model shows the type I error rates for factor A, factor B and the interaction effect. The type I error rates of the IPCW ANOVA method was close to the nominal rate of 0.05 at 25% censoring, and at all sample sizes for both the null and the main effects model. At 45% censoring, the type I error rates of the IPCW method was in general still close to the nominal rate of 0.05 although, it was a bit inflated in a few cases. The type I error rates of the AFT method was inflated when the number of observations in each cell before censoring was set at 10. At other cell sizes, it was close to the nominal rate of 0.05.

Table 6.3: Estimated mean and standard errors of the IPCW ANOVA and AFT when an interaction is present. n=50,σ^2=1

% censoring	Method	Estimate	MCSE	SE	Estimate	MCSE	SE	Estimate	MCSE	SE
		$\hat{\mu}_{11}$			$\hat{\mu}_{12}$			$\hat{\mu}_{21}$		
	TRUE	6.00			10.00			11.5		
25%	AFT	6.00	0.15		10.01	0.16		11.49	0.17	
	IPCW	6.00	0.16	0.16	10.01	0.16	0.16	11.49	0.17	0.16
	TRUE	6.00			10.00			11.5		
45%	AFT	6.00	0.17		10.02	0.18		11.49	0.20	
	IPCW	6.00	0.17	0.18	10.02	0.19	0.18	11.49	0.22	0.18
% censoring	Method	Estimate	MCSE	SE	Estimate	MCSE	SE	Estimate	MCSE	SE
		$\hat{\mu}_{22}$			$\hat{\mu}_{31}$			$\hat{\mu}_{32}$		
	TRUE	19.5			5			8		
25%	AFT	19.50	0.19		5.00	0.15		8.00	0.16	
	IPCW	19.49	0.20	0.16	5.00	0.15	0.16	8.00	0.16	0.16
	TRUE	19.5			5			8		
45%	AFT	19.51	0.49		5.00	0.16		8.00	0.17	
	IPCW	19.19	0.43	0.18	5.00	0.16	0.18	7.99	0.18	0.18

AFT=Accelerated Failure time Model, IPCW=IPCW ANOVA, MCSE=Monte Carlo SE

Table 6.4: Type I error rates and power of the IPCW 2×3 ANOVA model compared to the AFT model.

Cell size	% censoring	Method	Null Model			Main Effects Model			Interaction Model		
			Factor A	Factor B	Interaction	Factor A	Factor B	Interaction	Factor A	Factor B	Interaction
10	25%	AFT	0.054	0.066	0.064	1	0.877	0.067	-	-	0.834
		IPCW	0.045	0.056	0.054	0.999	0.833	0.055	-	-	0.744
	45%	AFT	0.047	0.050	0.082	0.997	0.801	0.089	-	-	0.62
		IPCW	0.048	0.062	0.052	0.978	0.653	0.070	-	-	0.266
20	25%	AFT	0.056	0.052	0.056	1	0.995	0.061	-	-	0.988
		IPCW	0.054	0.049	0.054	1	0.984	0.053	-	-	0.968
	45%	AFT	0.061	0.052	0.052	1	0.976	0.059	-	-	0.887
		IPCW	0.059	0.067	0.051	1	0.901	0.069	-	-	0.53
30	25%	AFT	0.044	0.061	0.055	1	0.999	0.059	-	-	0.999
		IPCW	0.047	0.059	0.054	1	0.999	0.048	-	-	0.999
	45%	AFT	0.046	0.052	0.055	1	0.995	0.051	-	-	0.983
		IPCW	0.056	0.053	0.065	1	0.975	0.070	-	-	0.773

IPCW=Inverse probability of censoring weighted ANOVA, AFT=accelerated failure time model

The main effects model shows the power of each test to detect the main effects and the type I error rate of the interaction model. Both the IPCW ANOVA model and the AFT model have greater than 80% power to detect the specified main effect at 25% censoring. The power to detect the specified main effect reduced for both models when the censoring rate was equal to 45%. The power of the IPCW method to detect the specified effect fell below 80% at small cell sizes (10) and large censoring (45%). In the presence of an interaction, the IPCW method did not acheive 80% power at small cell sizes (10) but had greater than 80% power to detect the interaction at 25% censoring and at greater cell sizes. In general, the IPCW method had lower power to detect the interaction term but performed better at larger cell sizes. In general, the IPCW model had comparable power to the AFT model at minimal censoring and large cell sizes but had lower power when either censoring was increased or the cell size was decreased.

Table 6.3 shows the estimates and standard errors of the IPCW ANOVA and AFT models. Both models produced consistent estimates of $\hat{\mu}_{11}, \hat{\mu}_{12}, \hat{\mu}_{21}, \hat{\mu}_{22}, \hat{\mu}_{31}, \hat{\mu}_{32}$. There was considerable bias for both models in the estimate of $\hat{\mu}_{22}$. This was reflected in the empirical standard errors of both models which was large compared to the standard errors of the other estimates. In the IPCW ANOVA model, the model based standard errors were comparable to their MC counterparts.

6.4 DISCUSSION OF IPCW ANOVA

A limitation of the Cox, AFT, Kaplan Meier and BJ methods is that they do not provide a means for decomposing the variation in the data. They also do not provide us with a measure of R^2 that has the simple interpretation as the proportion of varia-

tion explained. Furthermore, estimation methods based on iterations such as the BJ method may fail to converge to a reasonable estimate. The IPCW ANOVA method overcomes this limitation by providing reliable estimates of the mean parameters and an R^2 statistics that has the natural interpretation as the proportion of variation explained. The IPCW ANOVA method extends the simplicity and elegance of ANOVA to the censored data setting. The simulation studies in both the one way and two way IPCW ANOVA showed that although properties of the estimator where based on large sample theory, the model behaves well in small samples as long as censoring is minimal.

For many reasons including increased power and the test statistic being less sensitive to small departures from the homogeneity of variance assumption, balanced designs are preferred in the analysis of variance model. However, due to practical reasons such as censoring in the case of survival experiments, balanced designs may not always be possible. Whatever the cause, lack of balance necessitates care in the analysis of the data. If the data are only slightly unbalanced, there are several approximate procedures that might be used in the data analysis including randomly deleting some observations for cells containing a large amount of data in order to force balance into the data. For survival experiments, this is equivalent to deleting observations which are censored from that analysis. However, this reduces the precision of the estimates and even when censored observations are ignored, this does not ensure that the number of observations in each cell is equal since there is no guarantee that the number of censored observations will be equal in each cell. Further deleting observations to ensure balance results in an unnecessary waste of data. Moreover, different estimates of the model parameters will be obtained depending on which observations are further deleted.

As an alternative, for survival experiments, we may consider the censored observations as missing and use imputation methods for dealing with missing data. Here, one may replace censored values with alternate values estimated from the data. The most intuitive imputed value being the mean of the survival time. However, use of the unrestricted survival may tend to overestimate the survival times for these censored individuals. Furthermore, the use of the restricted mean may be problematic since the maximum survival time will likely be different for each cell. Moreover, it is well established that weighting is more advantageous than imputation for the handling of missing data.

The results of the one way and two-way inverse probability weighted analysis of variance show that this is a useful addition to the literature on extending the general linear model to the censored data setting when censoring is minimal. A key assumption of the method is that the data are missing at random. This assumption is more likely to be true when censoring is low. When this assumption is violated, which is likely to be true when the censoring is set at 45%, the model under performs the AFT model. The IPCW ANOVA model is an improvement over the Buckley James algorithm because it does not require computer intensive iterations and therefore does not run into convergences problems. The model is easy to use and readily implemented in available software.

Although tests based on the cell means including the inverse probability of censoring weighted analysis of variance test are generally more interpretable than those based on the effects model, the available statistical computing packages are designed for the effects model. Therefore care must be taken when implementing this method using standard available software, especially in the two factor case.

7.0 LENGTH BIASED RIGHT CENSORED SURVIVAL DATA

Length-Biased sampling is statistical artifact that occurs in survival analysis when the probability of an observation being included in the sample is proportional to a particular characteristic of that observation. It is often encountered in observational studies when the observed samples are not selected randomly from the population of interest, but with probability proportional to their length [8].

An example can be seen in the study of dementia and the onset of death in a Canadian Study of Health and Aging (CHSA) in 10,000 adults aged 65 and over [3]. Age of onset of dementia was obtained through care giver interviews and each patient was followed until death or the end of the study period. One of the goals of the study was to assess the effect of covariates on survival from the onset of dementia. The data is length biased because the cross sectional ascertainment of prevalent cases of dementia results in observed survival times that are longer than those from the population of interest because these individuals with dementia have to survive up to the time of recruitment to be included in the study.

Another example can be seen in the context of cancer screening. Slower-growing, less-aggressive tumors have a greater probability of detection than faster-growing more aggressive tumors because they have a longer pre-symptomatic period of time when they are detectable. Furthermore, these slower-growing cancers are

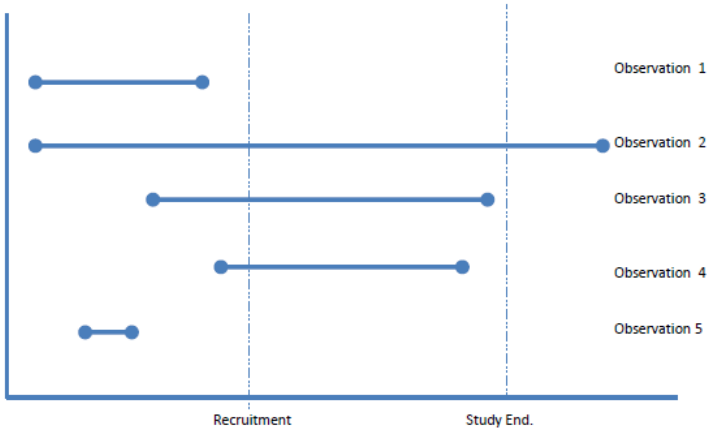

Figure 7.1: Example of a length-biased sample.

less likely to cause death. Therefore, patients with slower-growing cancers detected through a screening test will tend to have a longer survival period due to the indolent nature of the cancer. Thus length bias may artifactually make survival appear longer among patients who are detected in an early screening program compared to those detected clinically [13]. A schematic of a length biased sample is shown in Figure (7.1). In this diagram, Persons 1 and 5 experienced the event of interest but failed before the time of recruitment. Person 2 experienced the event of interest but their complete survival time is not fully known because of censoring. Persons 3 and 4 experienced the event of interest and their failure times are completely known. The data is length-biased because Persons 1 and 2 did not live long enough to be recruited into the study, therefore only those with longer survival times are sampled; the data is right censored because complete failure time is not completely know for Person 2.

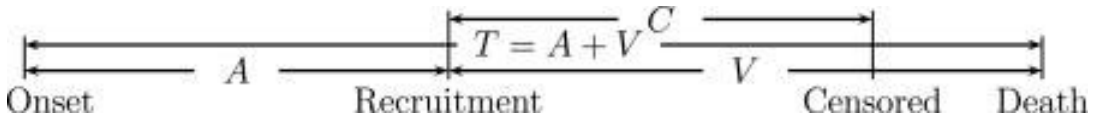

Figure 7.2: Schematic of a length-biased data.

In clinical studies, length biased sampling may occur when the disease in question is rare or due to the time and costs involved in sampling incident cases. An additional feature of these data is that the onset times including those not sampled are assumed to follow a stationary Poisson process. This assumption may be valid for some disease processes if the incidence rate of the disease is assumed to be constant over time. The effect of this sampling bias is that the sample mean is a biased estimator of the population mean [8].

Suppose the outcome of interest is a time to event. Let $\tilde{T}$ be the unbiased time measured from the initiating event to the terminating event having the density function $\tilde{f}(t)$ with mean $\tilde{\mu} = \int_0^\infty t\tilde{f}(t) < \infty$ and variance $\tilde{\sigma^2}$, A be the time from the initiating event to recruitment into the study and V be the residual time from recruitment to the terminating event. Then $\tilde{T} = A + V$. A is often referred to as the truncation times or the backward recurrence time and V is referred to as forward recurrence time or the residual lifetime. Also, let T be the corresponding length biased time measured from the initiating event to the terminating event with density $f(t)$. Figure (7.2) [47] is aids in understanding the definitions of these random variables. Under the stationarity assumption, that is assuming the truncation times A are uniformly distributed, the distribution of the length biased time T is

$$f(t) = \frac{t\tilde{f}(t)}{\tilde{\mu}}.$$

That is, the probability of inclusion in the sample is now proportional to the length of the survival time. If the unbiased survival time distribution is exponential, then the survival time distribution of the length biased sample will tend to be gamma with a shape parameter of 2. Thus the mean survival time of the length biased sample will be twice the mean survival time of an incident random sample drawn from the original population. Note that the k^{th} moment μ_k of the length biased distribution is

$$\begin{aligned} E_f(T^k) &= \int t^k f(t)dt \\ &= \int t^k \frac{t\tilde{f}(t)}{\tilde{\mu}} \\ &= \frac{1}{\tilde{\mu}} \int t^{k+1} \tilde{f}(t) \\ &= \frac{E(T^{k+1})}{\tilde{\mu}} \\ &= \frac{\tilde{\mu}_{k+1}}{\tilde{\mu}} \end{aligned}$$

Then when $k = 1$, $E_f(T) = \tilde{\mu} + \frac{\tilde{\sigma^2}}{\tilde{\mu}} > \tilde{\mu}$, which shows that the mean of the length-biased time is a biased estimator of the population mean.

Methodological contributions in the theory of length biased data in the absence of covariates include the work by Cox[8] who described the moments of the sampled lengths in terms of their unbiased distribution moments. Vardi[52][53] developed a nonparametric maximum likelihood estimator (NPMLE) of the true unbiased distribution function using the biased sample by attaching to each lifetime a weight which is inversely proportional to its size and gave conditions which guarantee the existence and uniqueness of the NPMLE.Large sample properties of the NPMLE were explored by Gill, Vardi and Wellner[17].

In the context of regression methods with biased responses, it is worth noting that the usual regression function does not hold for the biased responses since $E(T|X) = \tilde{\mu}(x) + \frac{\tilde{\sigma^2}(x)}{\tilde{\mu}(x)}$, where X is a covariate vector of interest. Hence direct application of regression methods lead to biased inferences. Skold[48], Cristobal and Alcala[10], Wu[57] and Cristobal et al[11] (among others) all have made significant contribution to the theory of regression methods for length biased responses. Wang[55] proposed an adaptation of the semi parametric proportional hazards regression model to estimate the effect of covariates on the length biased life times using a bias adjusted risk set in constructing the pseudo likelihood for the estimation of the model parameters. Chen[7] and Mandel and Ritov[33] proved the invariant property of the covariate effects in the AFT model for length biased outcomes and provided the estimating methods.

A more difficult situation arises when the outcome is not fully observed for some individuals. When length biased data is subject to right censoring, analysis of such data can be quite challenging due to the induced informative censoring since the survival time and censoring time are correlated through a common backward recurrence time. de Una-Alvarez[12] proposed a nonparametric estimator of the distribution function assuming independence between the survival and censoring times. Asgharian et al[3] overcame this stringent assumption of independence and assumed only independence between the residual lifetime measured from enrollment until the terminating event and the residual censoring time measured from enrollment. They adapted Vardi's NPMLE to the censored case.Linear regression type methods for length biased right censored data include the work by Qin and Shen[41] who proposed estimating equation based methods to assess the covariate effects under the semi parametric Cox model for length-biased data subject to right censoring.

Shen et al[47] used estimating equation methods to estimate covariate effects under the transformation model as well as the AFT model.

More recently Ning et al[39] have proposed a Buckley-James type to estimate the covariate effects for right censored length biased data using the model $log(\tilde{T}) = X^T\beta_0 + \epsilon$, where β_0 is a $p \times 1$ vector of covariates and ϵ has an unknown error distribution. In the presence of right censoring, the Buckley-James method replaces the censored observations with their conditional expectations. Ning et al [39] show that because the expectation of $\frac{logT_0}{T_0}$ is zero, where $T_0 = Texp(-x^T\beta_0)$, then a Buckley-James type estimator for the length-biased right-censored data is

$$U(\beta) = \sum_i^n \left\{ \delta_i \frac{logy_i - x_i^T\beta}{y_i exp(-x_i^T\beta)} + (1-\delta_i)\frac{\int_{y_{i0}}^{\infty} u^{-1} logu d\hat{F}_0(u;\beta)}{1-\hat{F}_0(y_{i0};\beta)} \right\},$$

where $y_i = \min(t_i, a_i + c_i), \delta_i = 1$ if failure occurs before censoring, and $\hat{F}_0(u;\beta)$ is an estimate of the unbiased survival function. The maximum likelihood estimate $\hat{\beta}$ of β does not have a closed form solution and is obtained through an iterative search

The goal of this study is to provide an alternative for the analysis of censored length biased life times under the accelerated failure time framework using inverse weighting techniques. A well known method for overcoming selection bias involves weighting each subject by their selection probabilities[20]. When the selection probabilities are known in advance, the weights can be treated as fixed in the analysis. However, in most practical research settings, these weights are not usually known in advance and need to be estimated from the data. The estimation of the weights needs to therefore be accounted for the resulting inference to be accurate. In the sections that follow, we propose a two stage weighted AFT model for the analysis of length biased right censored data in which the weights are estimated in the first stage and then used in the estimation of the regression parameters in the second

stage. Asymptotic properties of the parameter estimates are derived accounting for the variability introduced by the randomness in the weights and evaluated through simulation studies.

7.1 THE MODEL

Let $\tilde{T}$ be the unbiased failure time measured from the initiating event of the event until the terminating event, T be the length biased failure time measured from the initiating event of the event until the terminating event, A be the time from initiation to recruitment into the study, V be the residual life from recruitment until the terminating event, C^* is the unbiased censoring time measured from the initiating event until censoring and C be the residual censoring measured from recruitment until censoring. Then $T = A + V$ and $C^* = A + C$. T is only observed when $\tilde{T} > A$, which induces dependent censoring since $Cov(T, A + C) = Cov(A + V, A + C) = \sigma_A^2\{1 + \rho_{A,V}\sigma_V/\sigma_A\}$, where $\rho_{A,V}$ is the correlation between A and V and σ_V^2 = Var (V), similarly σ_A^2. An indicator $\delta_i = \begin{pmatrix} 1 & \text{if } T_i < C_i^* \text{ ;} \\ 0 & \text{otherwise.} \end{pmatrix}$ is used to differentiate a complete observation from a censored one. The complete data consists of $D_i = (Y_i, A_i, \delta_i, X_i)$ for each observation where $Y_i = \min(T_i, C_i*)$ and x_i is a p-vector of covariates. A usual assumption which holds true in many applications is the independence of the residual censoring time and the forward and backward recurrence times given the covariate vector i.e $\mathbf{C} \perp (\mathbf{A}, \mathbf{V})|\mathbf{X}$ since under stationarity the failure times and censoring time are positively correlated.

The model under consideration is the AFT model which relates the logarithm of the

failure time linearly to the covariates,

$$log(\tilde{T}) = \mathbf{X^T}\beta + \epsilon,$$

where ϵ has a specified distribution with mean zero.

7.2 ESTIMATION OF $\boldsymbol{\beta}$

In the absence of right censoring and biased sampling,the classical least squares principle provides an unbiased estimation equation for estimating the model parameter β. That is,

$$U(T, \beta|X) = \sum_{i=1}^{n} X_i(log(T_i) - X_i^T\beta) = 0$$

With length biased sampling, the probability of selection of an observation into the sample is directly proportional to its' survival time since only observations who have survived up to the time of recruitment can be sampled. The Horvitz-Thompson approach to correct for length bias is equivalent to weighting the observations by the inverse of their survival times. That is we can define the estimating function

$$U_{lb}(T, \beta|X) = \left\{ \frac{X_i(log(T_i) - X_i^T\beta)}{T_i} \right\}$$

which is unbiased for estimating β since

$$\begin{aligned} E\left\{\frac{X_i(log(T_i)-X_i^T\beta)}{T_i}\right\} &= \int \frac{X_i(log(T_i)-X_i^T\beta)}{T_i} \\ &= X_i\int \frac{log(T_i)-X_i^T\beta}{T_i}\frac{T_i\tilde{f}(t)}{\tilde{\mu}} \\ &= \frac{X_i}{\tilde{\mu}}E[(log(T_i)-X_i^T\beta)] \\ &= 0 \end{aligned}$$

Therefore, when length biased data is subject to right censoring, we propose the following estimating equation for estimating β

$$U_{lbc}(Y,\beta|X)=\sum_{i=1}^{n}\frac{1}{Y_i}\frac{\delta_i}{K_i(lnV|A)}X_i(log(Y_i)-X_i^T\beta)=0,$$

where $K_i(lnV|A) = P(lnC_i > lnV_i|A_i)$ is the conditional subject specific survival function of the censoring distribution. $U_{lbc}(Y,\beta|X)$ is an unbiased estimating equation for β since $h(\delta_i Y) = h(\delta_i T)$ and

$$\begin{aligned} E[\delta_i] &= E[E(\delta_i|A)] \\ &= \Pr(T_i < C_i^*|A) \\ &= \Pr(A_i + C_i > A_i + V_i|A_i) \\ &= \Pr(C_i > V_i|A_i) \\ &= \Pr(lnC_i > lnV_i|A_i) \\ &= K_i(lnV|A) \end{aligned}$$

Then

$$
\begin{aligned}
E(U_{lbc}(Y,\beta|X)) &= E\left(\frac{1}{Y_i}\frac{\delta_i}{K_i(lnV|A)}X_i(log(Y_i)-X_i^T\beta)|X\right)\\
&= E\left[\frac{1}{T_i}\frac{\delta_i}{K_i(lnV|A)}X_i(log(T_i)-X_i^T\beta)|X_i\right]\\
&= E\left[E\left[E\left(\frac{1}{T_i}\frac{\delta_i}{K_i(lnV|A)}X_i(log(T_i)-X_i^T\beta)\right)|T_i,A_i,X_i\right]\right]\\
&= E\left[E\left[\frac{1}{T_i}\frac{1}{K_i(lnV|A)}X_i(log(T_i)-X_i^T\beta)E(\delta_i|T_i,A_i,X_i)\right]\right]\\
&= E\left[E\left[\frac{1}{T_i}(log(T)-X^T\beta)\right]\right]\\
&= 0
\end{aligned}
$$

Therefore, an unbiased estimator for β is

$$\hat{\beta} = \left(\sum_i^n\left(\frac{\delta_i X_i X_i^T}{Y_i\hat{K}_i(lnV|A)}\right)\right)^{-1}\sum_i^n\frac{\delta_i X_i log(T_i)}{Y_i\hat{K}_i(lnV|A)}$$

where $\hat{K}_i(lnV|A)$ is a consistent estimator of $K_i(lnV|A)$ and is estimated using a parametric survival regression model to account for the dependence of the censoring and survival times.

Inverse probability weights can be both advantageous in counteracting bias due to unequal probabilities of inclusion among sampled units. However, when weights are disproportionately high, it introduces undesirable variability in the population statistics. Weight trimming is a common practice to reduce or diminish the effects of extreme weight values on the estimates and their estimated variances. Theoretically, a weight should be trimmed at the point where the loss of precision due to large weights is greater than the bias introduced by trimming these weights[40]. However, in practice ad hoc approaches that are not data driven are used to optimize the

variance/bias trade off. These ad hoc procedures are based on limiting the number and size of extreme weights by trimming or limiting the components used in the weight calculation or identifying and correcting for extreme weights after the weights have been calculated. The most common approach[40][2][26] is to trim weights larger than some value w_0 at w_0, where w_0 is typically chosen in an ad-hoc manner say at a constant value.

7.3 VARIANCE OF $\hat{\beta}$

Note that estimation of $\hat{\beta}$ requires estimation of $\hat{K}(.)$ from a model whose distributional assumption may rely on a parameter θ. Therefore, in order to assess the asymptotic variance of $\sqrt{n}(\hat{\beta}-\beta)$ that correctly reflects the estimation error of $\hat{K}(.)$, we consider the two step estimation as a one step M estimator

$$\sum_{i=1}^{n} h(D_i, \hat{\beta}, \hat{\theta}) = 0$$

where

$$h(D_i, \hat{\beta}, \hat{\theta}) = \begin{pmatrix} \psi(D_i, \beta, \theta) \\ \phi(D_i, \theta) \end{pmatrix}.$$

Then $V(\theta_0, \beta_0) = A(\theta_0, \beta_0)^{-1} B(\theta_0, \beta_0)\{(A(\theta_0, \beta_0))^{-1}\}^T$ where $A(\theta_0, \beta_0) = -E(\dot{h}(D_1, \theta_0, \beta_0))$ and $B(\theta_0, \beta_0) = E[h(D_1, \theta_0, \beta_0)h(D_1, \theta_0, \beta_0)^T]$.

For example, if the model $lnC_i = Z_i^T\gamma + \epsilon_i$ above is used in the first stage estimation of $\hat{K}(.)$, where Z_i is a p-vector including the backward recurrence time A and other

covariates, then the log likelihood function using an exponential survival model is

$$lnL = \sum_i^n \left\{ (1-\delta_i) ln\theta - \theta Y_i e^{Z_i^T \gamma} \right\}$$

so that the unbiased estimating equation for θ is given by

$$U(\theta) = \sum_i^n \frac{(1-\delta_i)}{\theta} - Y_i e^{Z_i^T \gamma}$$

and the estimate of the subject specific survival function for censoring is $\hat{K}_i(.) = e^{-\theta c exp^{Z_i^T \gamma}}$. Therefore the one step M estimator for β is given by:

$$U_3(\beta;\theta) = \sum_i^n h(D_i, \hat{\beta}, \hat{\theta}) = \begin{pmatrix} \sum_i^n \frac{\delta_i}{exp(-\theta Y_i e^{Z_i^T \gamma})} \frac{1}{Y_i} X_i (logY_i - X_i^T \beta) \\ \sum_i^n \frac{1-\delta_i}{\theta} - Y_i e^{Z_i^T \gamma} \end{pmatrix}$$

Then $\sqrt{n}(\hat{\beta} - \beta) \sim AN(0, \Sigma)$ where Σ is the upper left block of the variance covariance matrix $V(\beta_0, \theta_0)$. Specifically,

$$\begin{aligned} A(\theta_0, \beta_0) &= -E(\dot{h}(D_1, \theta_0, \beta_0)) \\ &= E \begin{bmatrix} \frac{\delta_i}{exp(-\theta Y_i e^{Z_i^T \gamma})} \frac{1}{y_i} X_i X_i^T & \frac{\delta_i}{Y_i e^{Z_i^T \gamma} exp(-\theta Y_i e^{Z_i^T \gamma})} \frac{1}{Y_i} X_i (logY_i - X_i^T \beta) \\ 0 & \frac{1-\delta_i}{\theta^2} \end{bmatrix} \\ &= \begin{bmatrix} E\left(\frac{X_i X_i^T}{y_i}\right) & E\left(\frac{\delta_i}{Y_i e^{Z_i^T \gamma} exp(-\theta Y_i e^{Z_i^T \gamma})} \frac{1}{Y_i} X_i (logY_i - X_i^T \beta)\right) \\ 0 & E\left(\frac{1-\delta_i}{\theta_0^2}\right) \end{bmatrix} \end{aligned}$$

and $B(\theta_0, \beta_0) = E[h(D_1, \theta_0, \beta_0) h(D_1, \theta_0, \beta_0)^T]$ where the entries of $B(\theta_0, \beta_0)$ are given

by

$$B_{11}(\theta_0;\beta_0) = E\left\{\left[\frac{\delta_i}{exp(-\theta Y_i e^{Z_i^T\gamma})}\frac{1}{Y_i}X_i(logY_i - X_i^T\beta)\right]\left[\frac{\delta_i}{exp(-\theta Y_i e^{Z_i^T\gamma})}\frac{1}{Y_i}X_i(logY_i - X_i^T\beta)\right]^T\right\}$$

$$B_{12}(\theta_0;\beta_0) = E\left\{\left[\frac{\delta_i}{exp(-\theta Y_i e^{Z_i^T\gamma})}\frac{1}{Y_i}X_i(logY_i - X_i^T\beta)\right]\left[\frac{1-\delta_i}{\theta} - Y_i e^{Z_i^T\gamma}\right]\right\}$$

$$B_{22}(\theta_0;\beta_0) = E\left\{\left[\frac{1-\delta_i}{\theta} - c_i e^{Z_i^T\gamma}\right]\left[\frac{1-\delta_i}{\theta} - Y_i e^{Z_i^T\gamma}\right]^T\right\}$$

Then $V(\theta;\beta) = A(\theta_0;\beta_0)^{-1}B(\theta_0;\beta_0)\{(A(\theta_0;\beta_0))^{-1}\}^T$. The variance parameters may be estimated from the data using their empirical moment estimators.

7.4 SIMULATION STUDY

We conducted simulations to assess the performance of the proposed estimator and compared it with the BJ-type estimator of Ning et al[39] and an inverse weighted length biased corrected estimator. In their paper, Ning et al use the beta estimates from the left truncation method by Lai and Ying [30] for general left truncated and right censored data as the initial values in the Buckley-James algorithm in the presence of covariate-dependent censoring. For practical reasons, we use instead the naive beta estimators from the linear model ignoring length bias and censoring as the initial starting values for the Ning et al BJ-type estimator to reduce the computation time since the method by Lai and Ying is itself computer. We used a 0.001 tolerance for all convergence loops and set the maximum number of iterations to 50. The estimates are based on 500 simulations.

Table 7.1: Comparison of the proposed estimator with the BJ-type estimator

Cohort Size	% censoring	Method	Estimate	MCSE	Model SE	Estimate	MCSE	Model SE	Estimate	MCSE	Model SE
				$\hat{\beta}_0$			$\hat{\beta}_1$			$\hat{\beta}_2$	
100	25%	TRUE	1.00			0.50			1.00		
		LB	0.19	0.67		0.65	0.64		0.93	0.96	
		BJ	0.92	0.19		0.52	0.13		1.09	0.21	
		LBC	1.04	0.14	0.16	0.51	0.11	0.10	0.98	0.18	0.37
	35%	TRUE	1.00			0.50			1.00		
		LB	0.05	0.73		0.57	0.73		0.87	1.05	
		BJ	0.91	0.23		0.51	0.16		1.12	0.25	
		LBC	1.03	0.16	0.12	0.51	0.12	0.08	1.00	0.21	0.26
200	25%	TRUE	1.00			0.50			1.00		
		LB	0.08	0.64		0.66	0.60		0.99	0.89	
		BJ	0.93	0.14		0.51	0.09		1.08	0.16	
		LBC	1.03	0.11	0.13	0.51	0.09	0.09	0.99	0.14	0.21
	35%	TRUE	1.00			0.50			1.00		
		LB	-0.03	0.63		0.57	0.64		0.91	0.93	
		BJ	0.94	0.17		0.52	0.11		1.10	0.19	
		LBC	1.03	0.11	0.13	0.51	0.09	0.06	1.00	0.15	0.16

LB=Length bias corrected, BJ=Buckley James type estimator, LBC=Length bias and censoring corrected

MCSE=Monte Carlo SE, Model SE= Model based SE

Table 7.2: Comparison of the proposed estimator under model misspecification

Cohort Size	% censoring	Method	Estimate	MCSE	Model SE	Estimate	MCSE	Model SE	Estimate	MCSE	Model SE
			$\hat{\beta}_0$			$\hat{\beta}_1$			$\hat{\beta}_2$		
		TRUE	1.00			0.50			1.00		
100	25%	LBC	1.04	0.14	0.13	0.51	0.11	0.08	0.98	0.18	0.22
	35%	LBC	0.19	0.11	0.07	0.51	0.08	0.04	0.99	0.14	0.08
		TRUE	1.00			0.50			1.00		
200	25%	LBC	0.19	0.16	0.11	0.51	0.12	0.07	1.01	0.21	0.19
	35%	LBC	0.19	0.12	0.07	0.51	0.09	0.04	1.02	0.15	0.09

LBC=Length bias and censoring corrected, MCSE=Monte Carlo SE, Model SE= Model based SE

We first generated independent pairs of $(A, \tilde{T})$. A was generated from a $Uniform(0, \tau)$ distribution where τ was larger than the upper bound of $\tilde{T}$ to ensure that the stationarity assumption held true. $\tilde{T}$ was generated using the AFT model $\log(\tilde{T}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$ where ϵ is a $N(0, 0.3)$ random variable, x_1 is Bernoulli with p=0.5 and x_2 is $Uniform(0, 1)$. To get the length biased sample we then selected pairs where $\tilde{T} > A$. Residual censoring times measured from enrollment were generated from a Weibull distribution with a shape of 1 and scale=$exp(\lambda * (a + x_1 + x_2))$. The value of λ was used to control the desired censoring rate at 0.25 and 0.35. We set $(\beta_0, \beta_1, \beta_2) = (1, 0.5, 1)$. A sample size of 100 and 200 was used to evaluate the estimator and the comparisons.

To estimate the robustness of the proposed estimator, we also assessed the performance of the proposed estimator under model misspecification. Here, data was generated according to the same scheme described above and the subject specific survival function of the censoring times was estimated using a lognormal distribution. Extreme weights were identified through empirical observation and corrected after the weight calculation process. For the proposed method, a maximum value of 5 was assigned to weights greater than 5.

Table 7.1 shows the results of the comparison of the proposed estimator with the Ning et al BJ-type estimator and the estimator accounting for length bias alone by inverse weighting. In general, compared with the BJ estimator and the proposed LBC estimator correcting for length bias and censoring, the LB estimator very poorly estimates the intercept $\hat{\beta}_0$ but performs slightly better in estimating $(\hat{\beta}_1, \hat{\beta}_2)$, although having considerably greater bias when compared with the BJ and LBC estimators. The BJ estimates have greater bias and less accuracy compared with the LBC estimator. The LBC estimator consistently estimates all model parameters. This is true across sample size and censoring. In general, the model based standard errors

of the LBC estimator are comparable to their Monte Carlo counterparts except at 25% censoring and at sample size of 100 where it is considerably greater. This could be as a result of a large number of extreme weights.

When censoring times are generated from a Weibull distribution but a Lognormal distribution is used in estimating the subject specific survival function, the results are presented in Table 7.2. Model standard errors are calculated in similar fashion as Table 7.1 using the Weibull AFT score function for comparison. The model parameters are still consistently estimated. However, the standard errors are smaller than their MC counterparts because they do not correctly reflect the variation in estimating $\hat{K}_i(.)$.

The simulation results suggest that the proposed estimator correcting for length bias and censoring is sufficient to achieve consistent estimates of model parameters when length biased data are subject to right censoring. Since the probability that an individual is censored is not known in advance and must be calculated from the data, this extra variation must be accounted for in the variance of the estimates to obtain valid inference. The AFT regression model is a desirable model for survival data due to the ease of interpretation. When length biased data are subject to censoring, the AFT model is useful for the linear regression analysis of the censored biased lifetimes. The Buckley James type estimator of Ning et al is a useful addition to the literature for censored regression of biased lifetimes. However, the algorithm is computer intensive and unlikely to be used in a real world setting. Furthermore, the asymptotic properties of the BJ-type estimator is dependent upon using a consistent estimator as the initial values in the iteration. In contrast, the proposed inverse weighted estimator is simple to use and easily implemented in standard available software. The asymptotic properties of the estimator established is not dependent on starting values.

7.5 REAL DATA EXAMPLE

The data used here is the Channing House from Hyde (1977) which is publicly available on the Klein and Moeschberger website. The data is on 462 persons in residence at the Channing House retirement home during the period of January 1964 to July 1975. The data consists of age at entry into home (in months), age at death or when the individual left the retirement home (in months), death status indicator and gender. Of interest is the lifetime of residents at death. The survival times of residents are length biased because a person must survive up to a given age to enter into the retirement community and as evidenced in Wang[54] Fig 3b, the stationarity assumption holds true. The data are also right censored because at the end of the observation period, not all persons in residence had died.

We applied the AFT model $E(log(Y)) = \alpha_0 + \alpha_1 gender$ to examine gender differences in the log survival time. Subject specific estimates of the survival function for the censoring distributed was estimated using a survival regression model assuming an extreme value distribution conditional on the truncation times. The estimated regression coefficients along with their standard errors are given in 7.3. The LBC estimates $(\hat{\alpha}_0, \hat{\alpha}_1)$ are comparable to the BJ estimates. The analysis shows that there is a significant difference in survival between males and females. On average, females residents at the residence home live longer than males. This inference is comparable to the analysis in Klein and Moeschberger[27] using the weighted Log rank test for the equality of the hazards between groups for left truncated and right censored data.

Table 7.3: Analysis of the Channing House Data

	LBC		BJ	LR Test
Parameter	Est	SE	Est	Inference
α_0	0.35	0.40	0.27	$\chi^2 = 1.82$
α_1	1.41	0.24	1.50	$p = 0.0341$

LBC=Length bias and censoring corrected

BJ=Buckley James type estimator, LR= Logrank test

7.6 DISCUSSION

Length-bias is a common sampling design bias occurring frequently in survival research. The bias arises because the probability that a life time is sampled is related to the length of the life time itself. The problem is further compounded when the failure times sampled are subject to censoring. Statistical analysis based on length biased samples have been the focus of much statistical research. When the data are length biased, the biased sampling will lead to biased inference on the population parameters unless the bias is corrected for the inference.

When length-biased data are subject to right censoring, the inference procedure must account for both length bias and censoring. If the measure of effectiveness is the survival time, length-bias as well as censoring results in the sample mean survival time being a biased estimator of the population mean. Length bias exaggerates the effectiveness, underscoring the need for proper analytic methods. In this dissertation, we considered a simple estimating equation approach for estimating model parameters when the data are both length-biased and right censored using inverse

weighting techniques. The simulation study shows that the proposed estimator has satisfactory finite sample performance. The simulation study and the data analysis example show that inversely weighting is sufficient to correct for both length bias and right censoring. Therefore, the mean survival may be correctly used as a measure of effectiveness for comparing between groups.

8.0 CONCLUDING REMARKS

All complete data survival analytic methods are based on assumptions. When data are missing, additional assumptions are required so that the observed data is the basis from which valid inferences are made. These additional assumptions that are imposed on the data should be as realistic and transparent as possible so that the observed data given the assumptions are still representative of the population. However, a problematic feature is that these assumptions cannot be verified from the data at hand. Sensitivity analysis where the robustness of the inference is examined is important to validate the inference. A key assumption of inverse probability weighting is that the data are missing at random. When this holds, inverse probability weighting recovers consistent estimates.

We have demonstrated the use of inverse probability weighting in the context of survival analysis of right censored, and length-biased right censored data. When the data are right censored, the simulations considered in this study use inverse probability techniques to extend analysis of variance to right censored data. A key feature of the proposed method is the focus on analyzing variability. The partitioning of variation is fundamental to experimental statistics. By examining variance components, we are able to understand our data in a way that we can not by simply looking at regression coefficients. Furthermore, when the effects of two treatments are to

be tested simultaneously, analytic methods for factorial designs with time-to-event outcomes have not fully been developed. The proposed method allows allows the interaction of treatments or the main effects of treatments by comparing the residual sums of squares between models as in the standard unbalanced analysis of variance. When data are both right censored and length biased, the literature on regression models for estimating model parameters is limited. The proposed method based on inverse weighting is a useful addition to the literature. The product of the inverse weights for the correction of length bias and censoring is adequate to eliminate the bias and provide consistent estimates of model parameters.

In both the right-censored, and right-censored length-biased setting, upon estimation of $\hat{K}_i(.)$ and $\hat{K}(.)$ by using the inverse of the product limit estimate of the survival function of the censoring distribution, the remaining inference is very straight forward and merges nicely into standard software packages. Hence, this approach can realistically be implemented by statistical practitioners.

The mean survival time is often used as a measure of effectiveness in screening and other public health programs. When the data are right censored and/or length-biased, the mean survival time is a biased estimator of the population mean. This work is of important public health significance because it provides an effective method of comparing health populations using the mean survival time as the measure of effect.

BIBLIOGRAPHY

[1] M. G. Akritas and M. P. LaValley. Nonparametric inference in factorial designs with censored data. *Biometrics*, 52:913–924, 1996.

[2] C.H. Alexander, S. Dahl, and L. Weidman. Making estimates from the american community survey. *Proceedings of the Social Statistics Section, American Statistical Association*, 2000, 1997.

[3] Masoud Asgharian, Cyr Emile M'Lan, and David B. Wolfson. Length-biased sampling with right censoring: An unconditional approach. *Journal of the American Statistical Association*, 97(457):201–209, 2002.

[4] N.E. Breslow. Covariance analysis of censored survival data. *Biometrics*, 30:89–99, 1974.

[5] B. W. Brown, M. Hollander, and R. M. Korwar. Nonparametric tests of independence for censored data, with applications to heart transplant studies. *Reliability and Biometry*, pages 327–354, 1974.

[6] Jonathan Buckley and Ian James. Linear regression with censored data. *Biometrika*, 66(3):429–436, 1979.

[7] Y.Q. Chen. Semiparametric regression in size-biased sampling. *Biometrics*, 66:149–158, 2010.

[8] D.R. Cox. Some sampling problems in technology. In *Developments in Survey Sampling*. Wiley, NewYork.

[9] D.R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society, Series B (Methodological)*, 34(2):187–220, 1972.

[10] J. A. Cristbal and J. T. Alcal. Nonparametric regression estimators for length biased data. *Journal of Statistical Planning and Inference*, 89:145168, 2000.

[11] J. A. Cristbal, J. L. Ojeda, and J. T. Alcal. Confidence bands in nonparametric regression with length-biased data. *Annals of the Institute of Statistical Mathematics*, 56:475496, 2004.

[12] J. de Ua-Alvarez. Product-limit estimation for length-biased censored data. *Test*, 11:109–126, 2002.

[13] Steven W. Duffy, Iris D. Nagtegaal, Matthew Wallis, Fay H. Cafferty, Nehmat Houssami, Jane Warwick, Prue C. Allgood, Olive Kearins, Nancy Tappenden, Emma O'Sullivan, and Gill Lawrence. Correcting for lead time and length bias in estimating the effect of screen detection on cancer survival. *American Journal of Epidemiology*, 168:98–104, 2008.

[14] B. Efron. The two sample problem with censored data. In *Proceedings, 5th Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press, Berkely, California.

[15] B. Efron. The efficiency of cox's likelihood function for censored data. *Journal of the American Statistical Association*, 72:557–565, 1977.

[16] R.D. Gill. Censoring and stochastic integrals. In *Mathematical Centre Tracts*. Mathematisch Centrum, Amsterdam.

[17] R.D. Gill, Y. Vardi, and J.A. Wellner. Large-sample theory of empirical distributions in biased sampling models. *The Annals of Statistics*, 16:1069–1112, 1988.

[18] Stephanie Green. Factorial designs with time-to-event endpoints. In *Handbook of Statistics in Clinical Oncology*, pages 199–210. Chapman and Hall.

[19] M. Greenwood. The natural duration of cancer. In *Reports on Public Health and Medical Subjects*, London. His Majesty's Stationary Office.

[20] D. G. Horvitz and D. J. Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260):663–685, 1952.

[21] P.J Huber. Robust estimation of location parameter. *Annals of Mathematical Statistics*, 35:73–101, 1964.

[22] J.O Irwin. The standard error of an estimate of expectation of life, with special reference to expectation of tumorless life in experiments with mice. *Journal of Hygeine*, 47:188–189, 1949.

[23] I.R. James and P.J. Smith. Consistency results for linear regression with censored data. *The Annals of Statistics*, 12(2):590–600, 1984.

[24] John D. Kalbfleisch and Ross L. Prentice. *The Statistical Analysis of Failure Time Data, Second Edition*. John Wiley Sons, Inc., Hoboken, New Jersey, 2002.

[25] E.L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282):457–481, 1958.

[26] L. Kish. Weighting for unequal pi. *Journal of Official Statistics*, 8, 1992.

[27] John P. Klein and Melvin L. Moeschberger. *Survival Analysis: Techniques for Censored and Truncated Data*. Springer, 2005.

[28] H. Koul, V. Susarla, and J. Van Ryzin. Regression analysis with randomly right-censored data. *The Annals of Statistics*, 9(6):1276–1288, 1981.

[29] Michael H. Kutner, Christopher J. Nachtsheim, John Neter, and William Li. *Applied Linear Statistical Models, Fifth Edition*. McGraw Hill Irwin, New York, New York, 2005.

[30] T. L. Lai and Z. Ying. Rank regression methods for left-truncated and right-censored data. *Annals of Statistics*, 19, 1991b.

[31] Tze Leung Lai and Zhiliang Ying. Large sample theory of a modified buckley-james estimator for regression analysis with censored data. *Annals of Statistics*, 19(3):1370–1402, 1991.

[32] Roderick J. A. Little and Donald B. Rubin. *Statistical analysis with missing data*. Wiley, New York, 2002.

[33] M Mandel and Y. Ritov. The accelerated failure time model under biased sampling. *Biometrics*, 66:1306–1308, 2010.

[34] Nathan Mantel. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotherapy Reports*, 50(3):163–170, 1966.

[35] Rupert Miller and Jerry Halpern. Regression with censored data. *Biometrika*, 69(3):521–531, 1982.

[36] Rupert G. Miller. Least squares regression with censored data. *Biometrika*, 63(3):449–464, 1976.

[37] N.J.D Nagelkerke. A note on a general definition of the coefficient of determination. *Biometrika*, 78:691–692, 1991.

[38] NIH. Aging interventions testing program, rfa: Ag-. 02-005.

[39] Jing Ning, Jing Qin, and Yu Shen. Buckleyjames-type estimator with right-censored and length-biased data. *Biometrics*, 67:1369–1378, 2011.

[40] F. Potter. A study of procedures to identify and trim extreme sample weights. *Proceedings of the Survey Research Methods Section, American Statistical Association*, pages 225–230, 1990.

[41] Jing Qin and Yu Shen. Statistical methods for analyzing right-censored length-biased data under cox model. *Biometrics*, 66:382–392, 2010.

[42] J. M. Robins and A. Rotnitzky. Recovery of information and adjustment for dependent censoring using surrogate markers. In *AIDS Epidemiology-Methodological Issues*, pages 846–866. N.Jewell, K.Dietz, and V.Farewell, Boston.

[43] James M. Robins, Miguel Angel Hernán, and Babette Brumback. Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11(5):550–560, 2000.

[44] Donald B. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701, 1974.

[45] M. Schemper and J. Stare. Explained variation in survival analysis. *Statistics in Medicine*, 15:1999–2012, 1996.

[46] S. R. Searle, G. Casella, and C. E. McCulloch. *Variance Components*. Wiley, New York, 1992.

[47] Y. Shen, J. Ning, and J. Qin. Analyzing length-biased data with semiparametric transformation and accelerated failure time models. *Journal of the American Statistical Association*, 104:1192–1202, 2009.

[48] M. Skld. Kernel regression in the presence of size-bias. *Journal of Nonparametric Statistics*, 12:41–51, 1999.

[49] E. Slud. Analysis of factorial survival experiments. *Biometrics*, 50:25–38, 1994.

[50] Peter J. Smith. *Analysis Of Failure and Survival Data*. Chapman Hall/CRC, Boca Raton, Florida, 2002.

[51] Leonard A. Stefanski and Dennis D. Boos. The calculus of m-estimation. *The American Statistician*, 56(1):29–38, 2002.

[52] Y. Vardi. Nonparametric estimation in the presence of length bias. *The Annals of Statistics*, 10:616–620, 1982.

[53] Y. Vardi. Empirical distributions in selection bias models. *The Annals of Statistics*, 13:178–205, 1985.

[54] M Wang. Nonparametric estimation from cross-sectional survival data. *JOurnal of the American Statistical Association*, 86, 1991.

[55] M. Wang. Hazards regression analysis for length-biased data. *Biometrika*, 83:343–354, 1996.

[56] L.J. Wei. The accelerated failure time model: A useful alternative to the cox regression model in survival analysis. *Statistics in Medicine*, 11(14-15):1871–1879, 1992.

[57] C. O. Wu. Local polynomial regression with selection biased data. *Statistica Sinica*, 10:789817, 2000.

[58] Hongwei Zhao and Anastasios A. Tsiatis. Efficient estimation of the distribution of quality-adjusted survival time. *Biometrics*, 55(4):1101–1107, 1999.