Bootstrapping and Survival Analysis

A Thesis Submitted to Dalhousie University In Partial Fulfillment of the Requirements for the Degree Bachelor of Science in Statistics With Honors

Fatma Sarhan

Supervisor: Dr. Edward Susko

April 24th, 2020

Contents

1 Abstract

This thesis will examine the effectiveness of bootstrapping on the confidence intervals of various parameters of interest. Analysis will be built up to see how bootstrapping behaves when it comes to the censoring issue in survival analysis. In order to do this, we start by looking at one samples (a single parameter) and then two samples (the ratio between two parameters) without and with censoring all from an exponential distribution. A real data analysis was also done to show a realistic depiction. Percentile and Percentile-t bootstrapping confidence intervals were calculated and compared with the Wald, which is a common method in survival analysis. A series of 1000 simulations were conducted, and we looked at the coverage percentage, width, lower and upper bounds to evaluate how well the methods performed. Some interesting results include, but are not limited to, seeing bootstrapping perform better with higher sample sizes and surprisingly with increased censoring.

2 Introduction

One of the interesting problems in survival analysis is estimating the unknown parameters that are included in the model that describe the lifetime of an individual in the study. We use X to represent the individual lifetime. There are different models that can be used to describe X. One of the commonly used lifetime models in survival analysis is the Weibull model. The probability density and survival functions for the Weibull model, respectively are

$$f(x;\theta) = \lambda\beta x^{\beta-1}e^{-\lambda x^{\beta}}\ ,\ x > 0\,, \tag{2.1}$$

and

$$S(x;\theta) = P(X > x;\theta) = e^{-\lambda x^{\beta}}\ ,\ x > 0\,, \tag{2.2}$$

where $\lambda > 0, \beta > 0$ are the model parameters. For simplicity, we use $\theta = (\lambda, \beta)$ to denote the vector of the model parameters.

This type of modeling is valuable in a real world setting when looking at the difference between two or more groups begin studied. For example, Leukemia data which entails two groups that have been treated with a different bone marrow are being analyzed with respect to their differences. To analyze this, we could look at the differences in the means of the Leukemia-free time between the two groups. However, the mean is not the only possible way of looking at the difference between the two but other entities such as the relative risk, interaction coefficients, etc...

Analysis will be built up, by first, making comparisons using a single sample from an exponential distribution with the point estimate of interest being λ. The comparisons will then be extended to two samples also from an exponential distribution with the point estimate of interest being $\frac{\lambda_1}{\lambda_0}$; without censoring. Further extension of this two sample study will be done through incorporating censored data, which is a special issue in survival analysis. It is essential to point out that the exponential distribution is a special case of the Weibull

model, mentioned above, where $\beta = 1$. The Wald test or confidence interval is the most common confidence interval construction procedure in survival analysis. On the other hand, bootstrapping methods provide a way of getting confidence intervals that have been shown, in some settings, to be robust to model assumption violations and that can have higher-order correctness properties. In this thesis, we will be testing the effectiveness of bootstrapping on the confidence intervals, of the above mentioned point estimates, using the Percentile and Percentile-t methods in comparison to the Wald test. At the end of this study, the Channing House real data will be analyzed for a real life representation.

3 Regression Models in Survival Analysis & Censoring

When it comes to any statistical analysis of any sort, it is important to note the notation used for random sampling, $x_1, ..., x_n$. This random sample is what is used to make inference about a parameter of interest regarding the population from which the sample was taken. The n serves as the size of this random sample and each x_i represents the value of interest observed for the ith observation. Specifically speaking, when it comes to survival analysis, the values of interest represent the lifetime of individuals, objects, etc...

Regarding survival data, there is a possibility that we run into censoring, which is a special problem arising in survival analysis. To elaborate, censoring occurs under the condition where a value is partially known. For instance, during a certain study lasting 5 years, patients going through chemotherapy are observed to see their lifetime progress during this period. Through the 5 years' time, some patients could pass away and others could live past the 5 years they were observed for. Those who pass away during the study time, we have an observed value for their lifetime notated regularly as x_i. On the other hand, patients that live past the study time, an exact lifetime is unknown, but rather we have partial information in that their lifetime was past those 5 years. Such an observation is considered censored and

is notated as $x_i{}^*$. The length of study or the time by which an individual leaves the study is annotated as c, and so a censored observation could be expressed as $x_i \geq c$; note that c could be fixed or random. Putting this all together, each observation is supposed to have a possibly unobserved failure time, x_i, and a censoring time, c_i. Explicitly speaking, any observed value during a study, where there is censoring involved, is observed as $t_i = \min(x_i, c_i)$. x_i and c_i are usually assumed independent. An additional indicator of failed vs censored observation, is the use of δ_i where $\delta_i = 0$ if the ith observation is censored and $\delta_i = 1$ if it has failed. With this said, the observations will be noted as (x_i, δ_i), where $i = 1, 2, \cdots, n$

An important note, for the purpose of this paper, focus will be on the Weibull model where $\beta = 1$; which yields an Exponential distribution. Given this, the likelihood equation is derived as follows:

$$L(\theta) = \prod_{i=1}^{n} [f(t_i; \theta)]^{I(\delta_i=1)} [S(t_i; \theta)]^{I(\delta_i=0)}, \tag{3.3}$$

where $I(A)$ is an indicator function defined as

$$I(A) = \begin{cases} 1, & \text{if } A \text{ is true,} \\ 0, & \text{otherwise.} \end{cases}$$

Substituting (2.1) and (2.2), when $\beta = 1$, into (3.3), the likelihood function for the exponential case is,

$$\begin{aligned} L(\lambda) &= \prod_{i=1}^{n} \left[\lambda e^{-\lambda t_i}\right]^{I(\delta_i=1)} \left[e^{-\lambda t_i}\right]^{I(\delta_i=0)} \\ &= \lambda^{\sum_{i=1}^{n} \delta_i} e^{-\lambda \sum_{i=1}^{n} t_i}. \end{aligned} \tag{3.4}$$

Using (3.4), the maximum likelihood estimate (MLE) will be derived and serves as an estimate of the parameter λ, denoted as $\hat{\lambda}$. The standard error (SE) of $\hat{\lambda}$ will be calculated

from the observed information matrix, which is also derived using the likelihood function. The way to which these entities are calculated will be outlined bellow.

The log-likelihood function is

$$\begin{aligned} l(\lambda) = \log(L(\lambda)) &= \log \lambda^{\sum_{i=1}^{n} \delta_i} + \log e^{-\lambda \sum_{i=1}^{n} t_i} \\ &= \left(\sum_{i=1}^{n} \delta_i \right) \log \lambda - \lambda \sum_{i=1}^{n} t_i. \end{aligned}$$

The first derivative of the log-likelihood function is

$$\frac{dl(\lambda)}{d\lambda} = \frac{\sum_{i=1}^{n} \delta_i}{\lambda} - \sum_{i=1}^{n} t_i.$$

To get the MLE of λ, we have to solve $\frac{dl(\lambda)}{d\lambda} = 0$. Doing so, the MLE becomes,

$$\hat{\lambda} = \frac{\sum_{i=1}^{n} \delta_i}{\sum_{i=1}^{n} t_i} = \frac{n^*}{n} \frac{1}{\bar{T}}, \tag{3.5}$$

where $n^* = \sum_{i=1}^{n} \delta_i$. Note: when there is no censoring $n^* = n$.

Now to get the SE, the observed information matrix needs to be calculated, $J(\theta)$. Generally, this information matrix is calculated as follows,

$$J(\theta) = -\frac{d^2 l(\theta)}{d^2 \theta}$$

For our purposes, the observed information matrix is arrived as shown bellow:
Using,

$$\frac{d^2 l(\lambda)}{d^2 \lambda} = -\frac{n^*}{\lambda^2},$$

therefore,

$$J(\lambda) = -(-\frac{n^*}{\lambda^2}) = \frac{n^*}{\lambda^2}.$$

From the above calculations, the SE is,

$$SE[\hat{\lambda}] = \sqrt{J(\hat{\lambda})^{-1}} = \sqrt{\frac{\hat{\lambda}^2}{n^*}}.$$

A part of the interest in this paper, as stated in the introduction, is looking at the Wald confidence interval and comparing it to the bootstrap confidence intervals of interest; will be mentioned later on. In order to calculate the Wald confidence interval, the aloft MLE of $\hat{\lambda}$ and its SE are used.

$$\hat{\lambda} \pm z_{\alpha/2} SE[\hat{\lambda}].$$

When it comes to survival regression models, the following log-linear model is most frequently used:

$$\log(x_i) = B_0 + B_1 z_i + \epsilon_i, \text{ where } x_i > 0 \text{ in a survival setting} \tag{3.6}$$

B_0 and B_1 are unknown constants (parameters) and z_i is a covariate. One of the differences with regards to the survival regression model, compared to the regular regression, is that the least squares are replaced using the likelihood method to adjust for the censoring. Additionally, the e^{ϵ_i} would follow a lifetime distribution. In this study, we will consider e^{ϵ_i} follows a Weibull or exponential distribution, instead of the usual case where $\epsilon_i \sim N(0, \sigma^2)$.

This regression model can be applied using the 'survreg' function in the R package to estimate the parameters of interest and their standard error. Note that this function uses the maximum likelihood methods like those mentioned above.

The regression model in (3.6) can be used to estimate λ. Under the assumption that e^{ϵ_i} follows $W(\lambda, 1)$, one can show that $B_0 = \log E[X] = -\log \lambda$ and therefore. In order to execute the log transformation, which will be analyzed in the One Sample section, we will

use the output from the 'survreg' function to first get the confidence interval of B_0.

$$\hat{B}_0 \pm z_{\alpha/2} SE(\hat{B}_0) = [L_1, U_1] \tag{3.7}$$

and then using this relationship, we can transform back to get the λ confidence interval after the transformation. This easily done as follows,

$$\left[e^{U_1}, e^{L_1}\right]$$

When working with the two sample case, without and with censoring, where one of the samples come from a population with parameter λ_0 ($W(\lambda_0, 1) = \text{exponential}(\lambda_0)$) and the other from a population with λ_1 ($W(\lambda_1, 1) = \text{exponential}(\lambda_1)$). The regression method was used to acquire the confidence interval of the ratio, $\frac{\lambda_1}{\lambda_0}$. In this situation, the z_i in the model is a factor that identifies which of the two samples the ith observation is from. Similar to before, $B_0 = -\log \lambda_0$ while $B_1 = -\log \lambda_1 + \log \lambda_0$. The ratio of $\frac{\lambda_1}{\lambda_0}$ is related to B_1 as follows,

$$e^{-B_1} = \frac{\lambda_1}{\lambda_0} \tag{3.8}$$

where B_1 is estimated by $\hat{B}_1$. Both $\hat{B}_1$ and $SE(\hat{B}_1)$ are taken from the 'survreg' output used. First, the confidence interval of $\hat{B}_1$ should be calculated as such,

$$\hat{B}_1 \pm z_{\alpha/2} SE(\hat{B}_1) = [L_2, U_2] \tag{3.9}$$

Then from the relationship noted in (3.8), we can get the ratios' confidence interval by exponentiating the lower and upper bound of $\hat{B}_1$ like so,

$$\left[e^{U_2}, e^{L_2}\right]$$

4 How does Bootstrapping Work?

Major methods of bootstrapping include both parametric and non-parametric methods. The advantage of the non-parametric methods is that there are no assumptions made on the distribution which can lead to more precise results. On the other hand, parametric methods can be easier to solve but may lead to incorrect results due to making assumptions about the distribution.

4.1 The Bootstrap Principle

Most generally, we have a homogeneous simple random sample $X_1, ..., X_n$ from a distribution whose cumulative distribution function (CDF) is denoted by $F(x;\theta)$. The given sample can be used to estimate the unknown parameter θ. Unfortunately, in several cases, it is very difficult to get the distribution of this parameters' estimate. As a result, we cannot calculate the confidence interval of the parameter. The bootstrap methods can be used to overcome this problem, as will be discussed later. This CDF can be estimated non-parametrically with what is known as the empirical distribution function, $\hat{F}$. Using a set of observation, $x_1, x_2, \ldots, x_n$, $\hat{F}$ is defined as the sample proportion given bellow

$$\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} I(x_i \leq x).$$

With large values of n, $\hat{F} \approx F$.

Consider a probability of interest where C is a fixed region and $g(\hat{\theta}, \theta)$ is a function of a random estimator $\hat{\theta}$ and the true but unknown θ. For instance, $g(\hat{\theta}, \theta) = \dfrac{(\hat{\theta} - \theta)}{SE(\hat{\theta})}$, which can be used to construct the confidence intervals.

Now, sampling from the above given sample we have, $X_1^*, ..., X_n^* \sim \hat{F}$. In the same way that $P_F(g(\hat{\theta}, \theta) \in C)$ can be calculated in theory, so can $P_{\hat{F}}((g(\hat{\theta}^*, \hat{\theta}) \in C)$; where $\hat{\theta}^*$ is the point estimate of θ from the bootstrap sample. What the bootstrap principle states is

that with large n, usually,

$$P_{\hat{F}}(g(\hat{\theta}^*, \hat{\theta}) \in C) \approx P_F(g(\hat{\theta}, \theta) \in C)$$

Since $P_{\hat{F}}(g(\hat{\theta}^*, \hat{\theta}) \in C)$, theoretically, does not have unknown values it can be calculated and used to get approximate confidence intervals; as will be demonstrated in subsequent sections.

4.2 Bootstrap Sampling

To begin, it is important to state that for the bootstrap confidence intervals, sampling with replacement from the the original sample was executed B times; total number of bootstrap samples. Using the bth bootstrap sample, $x_1^*, \ldots, x_n^*$, the estimate of θ, is denoted as $\hat{\theta}^{(b)}$, where $b = 1, \ldots, B$. This is equivalent to taking a simple random sample from $\hat{F}$. When $B = +\infty$, the corresponding distribution is that of $P_{\hat{F}}(g(\hat{\theta}^*, \hat{\theta}) \in C)$, which is the one of interest. It is important to point out that we cannot usually calculate $P_{\hat{F}}(g(\hat{\theta}^*, \hat{\theta}) \in C)$ in closed form. $g(\hat{\theta}^*, \hat{\theta}) \in C$ can be estimated through obtaining a large number of samples from $\hat{F}$ and then calculating the proportion of times $g(\hat{\theta}^*, \hat{\theta}) \in C$.

Why use bootstrap instead of large sample theory?

1. If we do not have a large sample (n) and can not assume the sampling distribution is normal.

2. If it is difficult to work out the standard error of the estimate.

4.3 Types of Bootstrapping Confidence Intervals

The distribution of $\hat{\theta}^{(b)}$ or $\hat{\theta}$ are generated using all the B bootstrap samples. Also, $c_{\frac{\alpha}{2}}$ and $c_{1-\frac{\alpha}{2}}$ are the $\frac{\alpha}{2}$ and $1 - \frac{\alpha}{2}$ quantiles of the $\hat{\theta}^{(b)}$ or $\hat{\theta}$ distribution respectively.

1. Percentile

The Percentile confidence interval is derived based on the distribution of $\hat{\theta}^{(b)}$ as

$$\left[c_{\frac{\alpha}{2}}, c_{1-\frac{\alpha}{2}} \right]$$

2. Percentile-t (Studentized)

   It is well known that

$$\hat{\theta} \sim \left(\theta_0, \frac{\hat{v}}{n} \right)$$

   The studentized confidence interval is derived based on the distribution of $\frac{\hat{\theta}^{(b)} - \hat{\theta}}{\sqrt{\frac{\hat{v}^{(b)}}{n}}} \simeq \frac{\hat{\theta} - \theta_0}{\sqrt{\frac{\hat{v}}{n}}}$

$$\left[\hat{\theta} - c_{(1-\frac{\alpha}{2})} \sqrt{\frac{v}{n}}, \hat{\theta} - c_{(\frac{\alpha}{2})} \sqrt{\frac{v}{n}} \right]$$

   The above confidence interval is derived through the following steps:

$$\begin{aligned} 1 - \alpha &= P \left(c_{\alpha/2} \leq \frac{\hat{\theta}^{(b)} - \hat{\theta}}{\sqrt{\frac{\hat{v}^{(b)}}{n}}} \leq c_{1-\frac{\alpha}{2}} \right) \\ &\simeq P \left(c_{\alpha/2} \leq \frac{\hat{\theta} - \theta_0}{\sqrt{\frac{\hat{v}}{n}}} \leq c_{1-\frac{\alpha}{2}} \right) \\ &= P \left(\hat{\theta} - c_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{v}}{n}} \leq \theta_0 \leq \hat{\theta} - c_{\frac{\alpha}{2}} \sqrt{\frac{\hat{v}}{n}} \right) \end{aligned}$$

   Note that, $\hat{v}$ is the variance from the original sample, while $\hat{v}^{(b)}$ is the variance from the bth bootstrap sample.

4.4 Higher Order Accuracy of Bootstrap Confidence Intervals

1. Percentile:

$$p[\theta_0 \epsilon CI] = 1 - \alpha + O(n^{-\frac{1}{2}}))$$

Where $O(n^{-\frac{1}{2}})$ means $p[\theta_0 \epsilon CI] = 1 - \alpha + R_n$

$$\left| \frac{R_n}{n^{-\frac{1}{2}}} \right| = \left| n^{\frac{1}{2}} R_n \right| \leq M(someM)$$

2. Percentile-t

$$p[\theta_0 \epsilon CI] = 1 - \alpha + O(n^{-1}))$$

Where $O(n^{-1})$ means $p[\theta_0 \epsilon CI] = 1 - \alpha + R_n$ [1]

$$\left| \frac{R_n}{n^{-1}} \right| = \left| n^{1} R_n \right| \leq M(someM)$$

4.5 Choice of Bootstrap Size

To make a choice regarding the value of B which works best, we perform a bootstrap on top of bootstrap. Meaning after sampling a bootstrap sample $(x_1{}^*, ..., x_n{}^*)$ with confidence interval represented as $[L, U]$, we resample from this bootstrap sample $(x_1{}^{**}, ..., x_n{}^{**})$. We do this resampling, from the bootstrap sample $(x_1{}^*, ..., x_n{}^*)$, B times and so we will have B confidence intervals represented as $[L^{(b)}, U^{(b)}]$; where $b = 1, 2, \ldots, B$. The standard error between all of these B resampled bootstrap samples is calculated as follows:

$$SE = \sqrt{\frac{\sum_{b=1}^{b=B} \left(L^{(b)} - \frac{\sum_{b=1}^{b=B} L^{(b)}}{b} \right)^2}{B - 1}}.$$

We expect the L from the original bootstrap sample to be within 2 SE of L with $B = +\infty$. The smaller the value of the SE corresponding to the B value used to calculate in relation to the confidence interval of the original sample $[L, U]$. Doing this method to choose the best B is more valuable if bootstrap is expensive. The cons to this method is that it is messy to implement. Although this was not done in this particular paper, it is a possible way of

[1] Davison and Hinkley, *Bootstrap Methods and their Application*, pg.39-40

making a choice regarding B. In principle, it is best to choose as large a B as possible. Yet since this is not feasible, we will point out how sensitive the results are to B through the simulations performed.

5 One Sample Results

This section will hold bootstrapping results based on working with one sample $x_1, ..., x_n$, with $n = 20$ first and then $n = 100$, from an exponential distribution with $\lambda = 1$. One thousand simulation were conducted, where every simulation calculations are based on a different randomly generated sample. Wald, Percentile, and Percentile-t confidence intervals for $\hat{\lambda} = \frac{1}{\bar{X}}$ are generated for all of these randomly generated samples. Analysis is done on these confidence intervals using the coverage percentage, the mean and variance of the confidence interval width, and the mean and variance of the lower and upper bounds.

The coverage percentage is an indicator of how many of the one thousand simulated confidence intervals include the true value of λ. The higher the value of the coverage percentage, the better due to this indicating more of the simulated confidence intervals actually including the value of interest. With respect to the mean width, the narrower the value of this the more improved the estimate. Since the range of possibilities for the estimate of λ will be less than that from a wider interval. Values of the mean of the lower and upper bounds that are closer to 1 are desirable. Finally, the lower the variance of the width, lower and upper bounds, the less variability exists between the simulated confidence intervals.

Prior to getting into the details, further clarification of how the 1000 simulations were used and what we are looking for. The simulations are used to approximate $P(\theta \in CI; \theta)$. Simply explained, for any given CI construction procedure, with 1000 simulations we have flipped a coin 1000 times and counted a head every time $\theta \in$ CI. The number of heads, X, is then binomial with parameters 1000 and $p = P(\theta \in \text{CI}; \theta)$. It follows that a 95% CI for

the true $P(\theta \in \text{CI}; \theta)$ is given by $\hat{p} \pm 2\sqrt{\frac{\hat{p}(1-\hat{p})}{1000}}$. Since the true $p \approx 0.95$, half the width of the interval is roughly $2\sqrt{\frac{0.95*0.05}{1000}} = 0.01378405$. In short, accounting for simulation error, the observed percentages, over the simulations, are likely to be within 1% or 2% of the true simulation error.

5.1 Tables of Results

The first set of two tables hold the coverage percentages for both $x_1, ..., x_{20}$ and $x_1, ..., x_{100}$ both coming from an exponential distribution with $\lambda = 1$. For the bootstrap confidence intervals, analysis was executed using 4 values of B; $B = 100, 1000, 5000, 10000$.

Table 1: Coverage percentages based on 1000 simulations at 4 different B values before the logarithmic transformation.

		n=20		
Wald	95.4%			
	$B = 100$	$B = 1,000$	$B = 5,000$	$B = 10,000$
Percentile	88.9%	90%	90.8%	90.5%
Percentile-t	89.3%	91%	91.3%	91.4%
		n=100		
Wald	95.9%			
	$B = 100$	$B = 1,000$	$B = 5,000$	$B = 10,000$
Percentile	93.8%	94.5%	94.9%	94.8%
Percentile-t	93.6%	94.6%	95.2%	95.4%

Table 2: Coverage percentages based on 1000 simulations at 4 different B values after the logarithmic transformation.

n=20				
Wald	94.2%			
	$B = 100$	$B = 1,000$	$B = 5,000$	$B = 10,000$
Percentile	88.7%	90.3%	90.6%	90.4%
Percentile-t	89.1%	90.9%	91.4%	91.6%
n=100				
Wald	95.7%			
	$B = 100$	$B = 1,000$	$B = 5,000$	$B = 10,000$
Percentile	93.7%	95%	95.1%	94.8%
Percentile-t	93.4%	94.9%	95%	95.5%

A quick overview of Tables 1 and 2, deeper analysis is done in the Comments section, conducting a log transformation of the parameters and doing bootstrapping with $B = 100$ did not seem to lead to better results. As a consequence, further analysis of the log transformation and $B = 100$ was not inquired.

Table 3: Mean and variance of the confidence interval widths based on the 1000 previously simulated samples at 3 different B values.

	$B = 1000$		$B = 5,000$		$B = 10,000$	
Method	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\mu}$	$\hat{\sigma}^2$
Wald	0.3949	0.0015	0.3949	0.0015	0.3949	0.0015
Percentile	0.3934	0.0031	0.3952	0.0030	0.3955	0.0030
Percentile-t	0.3856	0.0030	0.3875	0.0029	0.3874	0.0029

Table 4: Mean and variance of the confidence interval lower bounds based on the 1000 previously simulated samples at 3 different B values

	$B = 1000$		$B = 5,000$		$B = 10,000$	
Method	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\mu}$	$\hat{\sigma}^2$
Wald	0.8100	0.0063	0.8100	0.0063	0.8100	0.0063
Percentile	0.8401	0.0071	0.8393	0.0070	0.8392	0.0070
Percentile-t	0.8228	0.0069	0.8223	0.0068	0.8222	0.0068

Table 5: Mean and variance of the confidence interval upper bounds based on the 1000 previously simulated samples at 3 different B values

	$B = 1000$		$B = 5,000$		$B = 10,000$	
Method	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\mu}$	$\hat{\sigma}^2$
Wald	1.2048	0.0140	1.2048	0.0140	1.2048	0.0140
Percentile	1.2335	0.0155	1.2345	0.0153	1.2348	0.0153
Percentile-t	1.2085	0.0148	1.2099	0.0147	1.2096	0.0147

5.2 Comments

Coverage Percentage:

Based on both Table 1 and Table 2 coverage percentage results, we can see that with $n = 20$, the bootstrapping methods do not seem to have a very high coverage with values ranging from 88% to 91%; regardless the value of B. Increasing the sample size from $n = 20$ to $n = 100$, we can see that the bootstrapping has much with coverage percentages ranging from 93% to 96%. Additionally, we can see that the increased values of B also show significant increase in the coverage. With that said, the coverage values for $n = 20$ and $B = 100$, before and after transformation, show fairly low results $\approx 93\%$ compared to other values of B with

percentages $\approx 94\%, 95\%$. As a result, the cases where $n = 20$ and/or $B = 100$ are eliminated from further analysis.

Now looking at the affect of the log transformation, we see that doing this has not illustrated significant improvement in the coverage percentage for the bootstrapping confidence interval. For example, the coverage with $n = 100$ at $B = 10,000$ does not change and remains at 94.8% for the Percentile method. As for the Percentile-t, the increase in coverage is only by 0.1 from 95.4% to 95.5%. Contrarily, the Percentile-t results seem to be doing slightly better by rouphly a 1% increase for $B = 5,000$ and $B = 10,000$. However at $B = 100$ and $B = 1,000$ there is barely any difference in the coverage between the two methods.

In regards to comparison between the overall performance of the Percentile method versus the Wald test. Looking at the outcomes when $n = 20$, we can see that the Wald test is doing significantly better with a coverage percentage of 94% in contrast to 90.4% and 91.6% for the Percentile and Percentile-t respectively. When $n = 100$ we see that at $B = 100$ the bootstrap method yield values of 93.7% and 93.4%, which is slightly less better than a 95.7% coverage coming from the Wald test. Yet when we look at the other values of B we see that the bootstrap methods are doing much better with percentages ranging from 94.8%-95.5%. This range of values is still slightly lower than the 95.7%. However, it is important to note that, through increasing the sample size from 20 to 100 the difference between the Wald test and the bootstrap methods has decreased at values of B; with the exception of $B = 100$.

Finally, we will look at the differences between the two different bootstrap samples. Regardless of the value of n and B, we notice that the Percentile-t method always has a slightly higher coverage than the Percentile one. However, even though Percenile-t might be better, the difference between the two bootstrap methods is very little, with a maximum difference of 1%.

To further analyze the affects of the bootstrapping on the strength of the confidence intervals we will take a look at the confidence interval widths' mean and variance.

Confidence Interval Width mean and variance:

As mentioned previously continued analysis with $n = 20$, a log transformation and at $B = 100$ was removed due to the poor coverage percentage results illustrated. Looking at Table 3, which hosts the confidence interval widths' mean and variance with $n = 100$ and at 3 different B values, we can deduce the following observations. First, we will be commenting on the overall influence that the increase in B has on the mean and variance of the bootstrap methods. As the value of B increases for the Percentile method, the mean width shows very small increases of 0.0004 and 0.0021. The larger increase of 0.0021 occurs when going from $B = 1000$ to $B = 5000$ while the smaller increase is from $B = 5000$ to $B = 10000$. The difference within the Percentile-t method is very similar to that of the Percentile. An increase of 0.0019 in the mean width is seen when going from $B = 1000$ to $B = 5000$. Going from $B = 5000$ to $B = 10000$ however, and unlike in Percentile, we see a decrease of 0.0001 in the mean width. On the other hand, the variance of the width shows an opposite change to that of the mean. As the value of B increases from $B = 1000$ to $B = 5000$, and regardless of the bootstrap method, the variance decreases by 0.0001. Notice how it remains unchanged between $B = 5000$ and $B = 10000$.

Secondly, we will focus on the difference seen between the Wald and bootstrap method. When it comes to the mean of the width we see that there is an overall decreasing range of about $0.0006 - 0.0015$ and $0.0075 - 0.0093$ between Wald and Percentile and Wald and Percentile-t, respectively. So, with respect to the mean, the bootstrap methods provide narrower widths than that provided by the Wald. In contrast, the variance shows an increase of about $0.0015 - 0.0016$ and $0.0014 - 0.0015$, respectively. Hence, it is notable that there is a possible trade off between the mean and variance, although boostrapping provides narrower widths, it in turn has an increase in the variance; although very small.

Lastly, comments will be made on the difference seen between the bootstrap methods themselves. In general, it is apparent that the Percentile-t method provide both narrower

widths and smaller variances compared to the Percentile one. The decrease in the mean values, ranges from 0.0078 to 0.0081, while the variance decreases by 0.0001.

Lower Bound Mean and Variance:

As a reminder, when comparing lower bounds to one another, a higher value is better since that would be the closer to 1. As a start, let us look at the influence of increasing B on the mean and variance of the lower bound for the bootstrap methods. Most generally, as the value of B increases the mean seems to slowly decrease, while the variance decreases by a very small amount and then settles off. To elaborate, for the Percentile method, going from $B = 1,000$ to $B = 5,000$ the mean decreases by 0.0008 and from $B = 5,000$ to $B = 10,000$ by 0.0001. As for the Percentile-t, the mean value decreased by 0.0005 from $B = 1,000$ to $B = 5,000$ and by 0.0001 from $B = 5,000$ to $B = 10,000$. From this we notice that the amount of decrease in the mean value between $B = 1,000$ and $B = 5,000$ is larger than that from the decrease from $B = 5,000$ and $B = 10,000$; decreasing difference with increasing B. With respect to the variance, it is unlike the mean in terms of becoming consistent between $B = 5,000$ and $B = 10,000$ and decreases by 0.0001 from $B = 1,000$ to $B = 5,000$; such is the case with both of the bootstrap methods.

Regarding the difference between Wald and the bootstrap method, it is noticeable that the mean displays better values from the bootstrapping methods than the Wald. The bootstrapping method have a higher lower bound mean value than the Wald by at least 0.0128. The variance, contrarily, is lowest for the Wald with a value of 0.0063, and the maximum variance occurs for the Percentile with a value of 0.0070. So, the Wald is doing best with respect to the variance with a maximum difference of 0.0007 compared to the bootstrapping.

Finally, let us compare the two bootstrap methods to one another. Simply put, the Percentile showed the best performance when it comes to the mean compared to Percentile-t. Yet, Percentile-t shows smaller variance. Again, we see somewhat of a trade off between

the best mean values and the best variance when comparing any of the methods to one another.

Upper Bound Mean and Variance:

As a final attribute of analysis to test the affects of bootstrapping, we will look into the details of the upper bound mean and variance. Like the analysis of the lower bounds, we start off with studying the affect of B on the upper bound values. Recall that, lower values of the mean are considered better since they are closer to 1. As B increases, we see very small increases in the mean with the amount of increase being 0.0010 and 0.0003 for the Percentile. Notice that the amount of increase decreases from $B = 5,000$ to $B = 10,000$ with the value of 0.0003. For Percentile-t, there is an increase in the

Now, contrasting Wald to the bootstrapping methods, we see that the Wald is doing the best regarding both the mean and variance with values of 1.2048 and 0.0140, respectively. The difference between the Wald results and the maximum results from the bootstrap methods when it comes to the mean is 0.03. Quite a large difference relative to the difference we are used to seeing. So, the bootstrapping methods did not really result in the improvement of the upper bounds.

Our last observation is with regards to the difference between the two bootstrap ways. At a glance, it is clear that the Percentile-t method is doing significantly better than Percentile with smaller values of both mean and variance. The difference in means between the two range from 0.0246-0.0252, which is quite a large difference compared to what we have been seeing. The variance difference is 0.0007 (at $B = 1,000$) or 0.0006 (at both $B = 5,000$ and $B = 10,000$), and both methods display consistent variance values at the two high levels of B.

6 Two Sample Results

To further extend this study of the influence of bootstrapping method effectiveness, we will now study the confidence intervals of the ratio of two parameters from two different exponential samples. This analysis is more realistic for real life situations where we are interested in studying two or more groups. To elaborate, the bootstrapping results bellow are based on two exponential samples, each with a sample size of 100. One of the samples is from a population with parameter $\lambda_0 = 1$ and the other with parameter $\lambda_1 = 1.5$. With this said, the point estimate of interest is the ratio of $\frac{\lambda_1}{\lambda_0}$. A thousand simulations were generated from each of these two populations, and the Wald, Percentile, and Percentile-t confidence intervals were calculated by taking a pair from these 1000 paired samples; the first part of the pair is from the first population and the other from the second population. Similar to the analysis executed with the one sample above, we evaluate the effectiveness of the methods by looking at the coverage percentage, the width, and the lower and upper bounds. Note that, this two parameter investigation will be executed under two situations, without and with censoring.

6.1 Without Censoring

Coverage Percentage

These coverage percentage values look strange relative to what we expect. The values greatly fluctuate and are not consistent. As B increases the values go back and forth between 93% and 94%. Similarly, within the bootstrap methods, there is no consistent distinction between the two. This could be due to what is known as the Monte Carlo or simulation error.

Table 6: Coverage percentages based on 1000 simulations, each with a sample size of 100 at 3 different B values.

Method	$\mathbf{n = 100}$		
Wald	94.6%		
	$B = 1,000$	$B = 5,000$	$B = 10,000$
Percentile	94.1%	93.8%	94.3%
Percentile-t	94.2%	93.9%	93.9%

Confidence Interval Width

To start, we notice that the increase in the value of B results in an increase in the mean width, yet a decrease and stabilization in the variance. More explicitly, the mean increased by 0.0017 and 0.0002 for the Percentile and by 0.0039 and 0.00064 for the Percentile-t. Notice how the amount of increases decreases when B goes from 5,000 to 10,000; the high values of B. With respect to the variance, it decreases as B increases regardless the method. It decreases by 0.0007 (Percentile) and by 0.0003 (Percentile-t) when going from $B = 1,000$ to $B = 5,000$ and then stabilizes. The amount of increase is larger for mean in the Percentile-t and smaller for the variance.

Now, comparing the two methods together, we see that the Percentile-t method performs better than Percentile with regards to both the mean and the variance. The differences between the mean values are 0.0044, 0.0022, 0.0018. From this, we see that as B increases, the difference between the methods decreases. Variance differs between the two by either 0.0004 or 0 at $B = 1,000$, and $B = 5,000$ and $B = 10,000$, respectively.

Finally, looking at the difference between Wald and bootstrapping, we notice that the width of the better bootstrap method is narrower than that of the Wald. However, the variance of Wald is lower than that of the Percentile-t; again a trade off. The difference between the means of Wald and Percentile-t $0.0109 - 0.0154$, while the difference in variance

is either 0.0031 and 0.0034. Note that the difference has a range , since the bootstrap method differs at different B values.

Table 7: Mean and variance of the confidence interval width based on the 1000 previously simulated at 4 different B values.

	$B = 1000$		$B = 5,000$		$B = 10,000$	
Method	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\mu}$	$\hat{\sigma}^2$
Wald	0.8487	0.0146	0.8487	0.0146	0.8487	0.0146
Percentile	0.8377	0.0184	0.8394	0.0177	0.8396	0.0177
Percentile-t	0.8333	0.0180	0.8372	0.0177	0.8378	0.0177

Confidence Interval Lower Bound

Just like the case of the mean width, the lower bound values decreases in correspondence to the increasing B values; getting worse. The variance shows little to know variation depending on the method. In the Percentile method, the mean value decreases by 0.0009 an 0.0002 and the variance remains unchanged relative to B. Notice again the amount of difference as B becomes bigger, decreases. Similarly, the Percentile-t method shows difference of 0.002 and 0.0003, while the variance decreases by 0.0005 when going from $B = 1,000$ to $B = 5,000$ and remains the same.

Now, looking between the bootstrap ways, we notice that Percentile-t is doing better with mean values that are always higher than those of the Percentile. The difference between them ranges from $0.0003 - 0.0008$. The variance however does not illustrate a consistent trend throughout, there are time at which the variance is lower for Percentile and other times vice versa. It is important to note however, that the Percentile variance is doing better by 0.0004 only at $B = 1,000$ and then not so at higher Bs with a smaller difference of 0.0001.

Wald is doing better than the bootstrap methods when the bootstraps is at the two highest levels of B. To elaborate, the only time at which the bootstrap performs better

than the Wald, when it comes to the mean, is at $B = 1,000$. Regarding the variance, Wald is continuously better with the smallest variance of 0.0026 compared to the other values of 0.0273, 0.0274, or 0.0278.

Table 8: Mean and variance of the confidence interval lower bound based on the 1000 previously simulated at 4 different B values

Method	$B = 1000$		$B = 5,000$		$B = 10,000$	
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\mu}$	$\hat{\sigma}^2$
Wald	1.1456	0.0266	1.1456	0.0266	1.1456	0.0266
Percentile	1.1502	0.0274	1.1493	0.0274	1.1495	0.0274
Percentile-t	1.1522	0.0278	1.1501	0.0273	1.1498	0.0273

Confidence Interval Upper Bound

Once again, increasing values of B, correspond to increasing mean values of the upper bound; which is worse. The mean increases by 0.0006 and 0.0005 for Percentile, and by 0.018 and 0.0003 for Percentile-t. Yet again, there is a decrease in the difference as B increases. With respect to the variance, it decreases steadily as B increases.

Overall, there is no significantly better method since the difference between the methods' values are very slight. Regardless of this, the Percentile-t has both better mean and variance values than Percentile; as seen previously. The difference in mean between the two ranges from $0.0013 - 0.0025$. On the contrary, variance does not show a strict pattern distinguishing a method over the other. At $B = 1,000$, Percentile-t is best by 0.0003, at $B = 5,000$ they are the same, and at $B = 10,000$ the Percentile-t once again does better but by 0.0002 now. It is possible that the 0 difference between the two ways is an exceptionaly case, given that Percentile-t does better once again at a higher B.

Comparing Wald to the bootstrap, we see that once more there is a trade off between the better mean value and the variance. The mean values of the bootstrap methods show better

results with differences ranging from $0.0052 - 0.0088$. The variance, perhaps as would be expected due to consistency in previous analysis, is the smallest compared to all bootstrap methods with difference ranging from $0.0007 - 0.0012$.

Table 9: Mean and variance of the confidence interval upper bound based on the 1000 previously simulated at 4 different B values

	$B = 1000$		$B = 5,000$		$B = 10,000$	
Method	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\mu}$	$\hat{\sigma}^2$
Wald	1.9943	0.0806	1.9943	0.0806	1.9943	0.0806
Percentile	1.9880	0.0818	1.9886	0.0814	1.9891	0.0815
Percentile-t	1.9855	0.0815	1.9873	0.0814	1.9876	0.0813

Final Overall Comment

With increasing values of B (from 1,000 to 10,000), specific entities might not be necessarily doing better in terms of wanted values. Yet, it is important to note that we consistently see the differences decrease between the entity values (mean width, lower bound, and upper bound) as B increases. This perhaps, illustrates that although increasing B does not provide improved estimates, there is a stabilization of the values shown in the decreased differences.

6.2 With Censoring

6.2.1 Dealing with Censoring When Bootstrapping

To deal with the censoring evident in this case, the only difference is the the 'survreg' function used. Unlike before, we have incorporated censored values in the setup of the function. From here on the same procedure is done, as explained at the end of the third section of this paper.

6.2.2 Fixed Censoring

In order to test the bootstrapping affects on the estimation accuracy of parameters of interest when it comes to censoring, fixed censoring was done for simplicity. Three fixed censoring values were calculated based on three corresponding proportions, $p = 0.10, 0.20,$ and 0.50, where p is the proportion of censored values in a sample; recall $p = \dfrac{n - n^*}{n}$. The following steps show how the censoring values were calculated:

Note: Given our overall sample is split in half, 50% from a population an exponential population with λ_0 and 50% from a population with parameter λ_1. Since we want p of this overall sample to be censored, then half of this p will come from population 0 and the other half from population 1. Thus the sum is expressed as follows.

$$\frac{1}{2}e^{-\lambda_0 c} + \frac{1}{2}e^{-\lambda_1 c} = p, \text{ where } \lambda_0 = 1 \text{ and } \lambda_1 = 1.5$$

$$e^{-c} + e^{-1.5c} = 2p$$

If $e^{-1.5c} > p$, then $e^{-c} + e^{-1.5c} > 2p$ as well, and so c would be too small since $c < \frac{-\log(p)}{1.5}$. If however, $e^{-c} < p$, then $e^{-c} + e^{-1.5c} < 2p$ and c would then be too large since $c > -\log(p)$. As a result,

$$\frac{-\log(p)}{1.5} < c < -\log(p)$$

Therefore the best approximate of c would be the minimum of the set of values that satisfy $e^{-c} + e^{-1.5c} < 2p$. Now, when going through the above steps with the three different p values, we get $c_1 = 1.9323$, $c_2 = 1.3313$, and $c_3 = 0.5625$.

6.2.3 Results

Coverage Percentage

In order to take a look at the affect of bootstrapping when it comes to censoring, we will specifically look at the affect as censoring increases in addition to the same points of analysis as before. To start, we will look at the overall coverage percentages as the amount of censoring increases. From the following Table 10, we can see that there is a continuous slight increase in the coverage as c_r decreases (p increases). For the bootstrapping methods, at c_1 we have about 93% and low 94%, then slowly increase at c_2 to higher 94%, and then we go up to 95% at c_3. Wald on the other just hovers at about 94%/95% coverage regardless of the censoring.

Looking specifically at the way B influences the coverage when censoring is involved, we see that there is slight increases ranging from 0.1 - 0.3 and then a steadiness at the two higher B values. However, it is important to note that there were three exceptions to this. The Percentile-t at c_1, the coverage goes up and down by 0.1, which does not exactly illustrate the increase and steadiness mentioned above. Similarly, at c_3 we see that same fluctuation. Yet, such small fluctuations are not significant and could be due to what has been previously introduced as the Monte Carlo or simulation error. In general, it does not seem that increases in B show more steady or improved results with the increased censoring.

Now, regarding which bootstrap method shows better results, it seems that mostly Percentile is doing best. This is seen at both c_1 and c_3, with the exception of c_2 where Percentile-t shows better performance. This is somewhat different than what we have seen before, without censoring, where Percentile-t illustrated better results; in accordance with the theory. Now comparing the Wald with the bootstrap, we see that the Wald is doing better yet as the censoring increases the gape between the bootstrap methods and the Wald decreases. Hence, from this we can say that the censoring seemed to bridge the gap between the bootstrap methods and the Wald. Yet, it also caused strange results by showing better values for the

Percentile than the Percentile-t as apposed to what we expect.

Table 10: Coverage percentages based on 1000 simulations, each with a sample size of 100 at 3 different B values.

c_1 = 1.9323			
Wald	94.5%		
	$B = 1,000$	$B = 5,000$	$B = 10,000$
Percentile	93.9%	94.2%	94.2%
Percentile-t	94.0%	93.8%	94.1%
c_2 = 1.3313			
Wald	94.3%		
	$B = 1,000$	$B = 5,000$	$B = 10,000$
Percentile	94.2%	94.4%	94.4%
Percentile-t	94.4%	94.5%	94.5%
c_3 = 0.5625			
Wald	95.4%		
	$B = 1,000$	$B = 5,000$	$B = 10,000$
Percentile	95.1%	95.3%	95.2%
Percentile-t	95.0%	94.9%	95.0%

Confidence Interval Width

From Table 11, it is inevitable that with increased censoring values, there is an increase in both the mean width and the variance of the confidence intervals. The largest mean width value is 1.2877 at c_3 and $B = 10,000$, while the lowest value is 0.8935 at c_1 and $B = 1,000$, showing the maximum difference between the mean width being 0.3942; quite a large difference. The variance on the other hand, is largest at c_3 and $B = 1,000$ with a value of 0.0907 and smallest at c_1 with value of 0.0199; difference being 0.0708.

The increase in B shows an overall similar behavior to the results received in the without censoring, and that is regardless of the censoring level. To elaborate, as B increases, we see that the mean width increases while the variance decreases; just like before. Hence, involving censoring did not result in different findings. Similarly, Percentile-t exhibits better values with respect to both the mean and variance compared to the Percentile values. However, recall that previously both bootstrapping methods were performing better with respect to the mean width, while the Wald had lower variance; the trade off. With censoring, the Wald seems to be doing worse only in comparison to the Percentile method in regards to the mean; unlike before.

Table 11: Mean and variance of the confidence interval width based on the 1000, each with a sample size of 100 at 3 different B values.

$c_1 = 1.9323$						
Wald	$\hat{\mu} = 0.9018$		$\hat{\sigma}^2 = 0.0199$			
	$B = 1,000$		$B = 5,000$		$B = 10,000$	
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\mu}$	$\hat{\sigma}^2$
Percentile	0.9023	0.0235	0.9052	0.0230	0.9057	0.0229
Percentile-t	0.8935	0.0224	0.8979	0.0223	0.8986	0.0223
$c_2 = 1.3313$						
Wald	$\hat{\mu} = 0.9621$		$\hat{\sigma}^2 = 0.0262$			
	$B = 1,000$		$B = 5,000$		$B = 10,000$	
Percentile	0.9671	0.0303	0.9706	0.0295	0.9711	0.0295
Percentile-t	0.9557	0.0287	0.9607	0.0286	0.9616	0.0283
$c_3 = 0.5625$						
Wald	$\hat{\mu} = 1.2571$		$\hat{\sigma}^2 = 0.0796$			
	$B = 1,000$		$B = 5,000$		$B = 10,000$	
Percentile	1.2814	0.0907	1.2871	0.0897	1.2877	0.0889
Percentile-t	1.2502	0.0813	1.2550	0.0806	1.2560	0.0809

Confidence Interval Lower Bound

In the same way that the mean width and variance worsened with higher censoring, poorer lower bound values are seen with the increased censoring. More explicitly, the lower bound decreases as we transition from c_1 to c_2 to c_3, which is undesired since we want higher lower bound values. Correspondingly the variance also increases; again undesired.

Similar to before, the variance continuous decreases with increasing B, yet the mean does not show similar distinct results as seen prior. At times, the mean decreases, and in the case of c_1 for Percentile the mean increases, and a lot of times there is no specific trend, just fluctuating. When it comes to differences between the bootstraps, the Percentile-t is not doing best in most cases like before, it is only doing better because of the continuous lower variance values. Finally, comparing Wald with the bootstraps we see that the bootstrap is doing better for the mean only at c_1 and c_2 but not so at c_3, while the variance is lower for Wald; similar to without censoring results. At c_3, there are not distinct results, as there is a lot of fluctuations.

Table 12: Mean and variance of the confidence interval lower bound based on the 1000, each with a sample size of 100 at 3 different B values.

c_1 = 1.9323						
Wald	$\hat{\mu} = 1.1325$		$\hat{\sigma}^2 = 0.0291$			
	$B = 1,000$		$B = 5,000$		$B = 10,000$	
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\mu}$	$\hat{\sigma}^2$
Percentile	1.1369	0.0301	1.3557	0.0300	1.3571	0.0299
Percentile-t	1.1356	0.0299	1.3375	0.0294	1.1332	0.0292
c_2 = 1.3313						
Wald	$\hat{\mu} = 1.1118$		$\hat{\sigma}^2 = 0.0312$			
	$B = 1,000$		$B = 5,000$		$B = 10,000$	
Percentile	1.1154	0.0324	1.1148	0.0323	1.1147	0.0322
Percentile-t	1.1148	0.0317	1.1128	0.0315	1.1124	0.0314
c_3 = 0.5625						
Wald	$\hat{\mu} = 1.0271$		$\hat{\sigma}^2 = 0.0425$			
	$B = 1,000$		$B = 5,000$		$B = 10,000$	
Percentile	1.0328	0.0451	1.0271	0.0425	1.0312	0.0448
Percentile-t	1.0329	0.0441	1.0308	0.0446	1.0309	0.0435

Confidence Interval Upper Bound

Once again we see that the censoring results in the values of the upper bound to become worse. That is since with increased censoring the values of the mean and variance are increasing which is contrary to what we want.

With increased B, the upper bound values are increasing while the variance is decreases. This is, again, similar to our conclusion without censoring involved. Additionally, the Percentile-t is performing better with smaller values for the upper bound mean and variance. Now, comparing these bootstraps to the Wald, we can see that only the Percentile-t

is continuously doing better with respect to the mean, unlike the Percentile. The variance however, is always best from the Wald.

Table 13: Mean and variance of the confidence interval upper bound based on the 1000, each with a sample size of 100 at 3 different B values.

$c_1 = 1.9323$						
Wald	$\hat{\mu} = 2.0342$		$\hat{\sigma}^2 = 0.0968$			
	$B = 1,000$		$B = 5,000$		$B = 10,000$	
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\mu}$	$\hat{\sigma}^2$
Percentile	2.0392	0.0999	2.0408	0.1001	2.0414	0.1000
Percentile-t	2.0291	0.0976	2.0317	0.0976	2.0318	0.0974
$c_2 = 1.3313$						
Wald	$\hat{\mu} = 2.0740$		$\hat{\sigma}^2 = 0.1141$			
	$B = 1,000$		$B = 5,000$		$B = 10,000$	
Percentile	2.0825	0.1193	2.0854	0.1189	2.0859	0.1190
Percentile-t	2.0705	0.1153	2.0736	0.1157	2.0740	0.1152
$c_3 = 0.5625$						
Wald	$\hat{\mu} = 2.2841$		$\hat{\sigma}^2 = 0.2348$			
	$B = 1,000$		$B = 5,000$		$B = 10,000$	
Percentile	2.3143	0.2559	2.3179	0.2547	2.3189	0.2542
Percentile-t	2.2830	0.2379	2.2862	0.2375	2.2869	0.2377

7 Real Life Data Analysis

In this part, we will be analyzing the Channing House Data. This data set consists of 462 individuals in a retirement community and we are interested in whether there is a difference in the time of death between men and women. The data consists of 7 columns which are death status, the age of entry in the retirement home, age of exit or death from

the retirement home, the time difference between the above two ages, and the gender. The relative risk is the point estimate of interest and is the value that shows the difference in time of death between the two genders. We will fit the data to both Weibull and Log-logistic models. Comparison will be done based on the Wald, Percentile and Percentile-t confidence intervals obtained from the two regression outputs, the maximized log likelihood values, and the deviance residuals. Similar to previous analysis, the bootstrap intervals will be calculated using $B = 1,000$, $B = 5,000$, and $B = 10,000$. Note that, this data set consists of not only right censored observations, but also left truncated ones. The 'survreg' R function, which is used in this analysis, does not allow taking into account the left truncated individuals, and hence the truncation was ignored.

To elaborate, relative risk is $\frac{\zeta_z}{\zeta_0} = e^{\mathbf{z}^T\gamma}$; relative to baseline. In our case, the baseline is set as the women, and the outputed relative risk values are $e^{-0.3436} = 0.7399$ from Weibull and $e^{-0.3436} = 0.7092$ from Log-logistic. Theses values imply that lifetime is expected to be shorter for men than that for women in this study.

To start, Table 14 exhibits the confidence interval results for both models.

Table 14: Confidence Intervals for the Channing House data set when the data is fitted using both Weibull and Log-logistic

Weibull			
Wald	[0.5818, 0.9411]		
	$B = 1,000$	$B = 5,000$	$B = 10,000$
Percentile	[0.5798, 0.9457]	[0.5774, 0.9494]	[0.5800, 0.9431]
Percentile-t	[0.7007, 1.0301]	[0.6984, 1.1292]	[0.6983, 1.1069]
Log-logistic			
Wald	[0.5371, 0.9365]		
Percentile	[0.5374, 0.9406]	[0.5379, 0.9356]	[0.5407, 0.9370]
Percentile-t	[0.6649, 1.0698]	[0.6640, 1.1444]	[0.6639, 1.1281]

From the above Weibull confidence interval results, in Table 14, we can see that the confidence interval widths range from 0.3594 to 0.4308. On the other hand, we have the widths for the Log-logistic ranging from 0.3963 to 0.4804. From this, we can see that the Log-logistic confidence intervals tend to have slightly wider intervals; recall that we desire narrower confidence intervals. Thus, based on this analysis of the widths of the confidence intervals, we can expect the Weibull model to be a better fit for this data.

Another way of determining the best fit model is by looking at the values of the maximized log likelihood for each of the distributions from the summary output of the regression model. The distribution which yields the larger log likelihood value is chosen as the best option. In doing so, Weibull had a value of -1102.82 compared to that of -1105.10 from the Log-logistic model; the difference is small but it is still there. Hence, based on the log likelihood values, Weibull is a slightly better fit for the data as previously anticipated.

Finally, we will take a look at the deviance residuals in order to assess the more appropriate functional form for the Channing House data. From the below Figure 1, we can see that the deviance residuals from the Weibull model have an average that is closer to the 0 horizontal line for the males (gender = 1); which is desired for residuals. The average residuals for the females (gender = 2) however, does not show much differences between the two models. Additionally, the Weibull plot does not have any outliers at all in comparison to the Log-logistic. When it comes to the IQR, both genders in both models do not show great differences from one another, although the Weibull shows very slightly larger IQR than Log-logistic. Therefore, once again the Weibull model shows slightly better performance than the Log-logistic one.

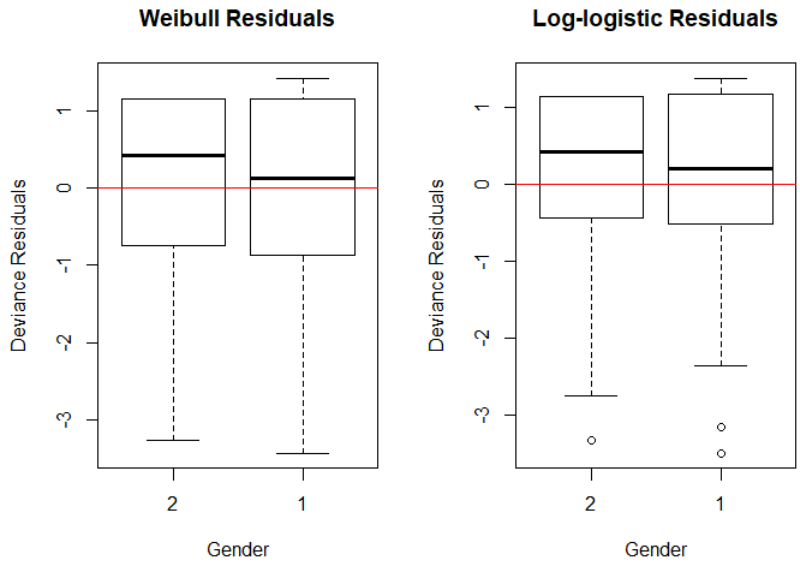

Figure 1: The deviance residuals of both the Weibull and Log-logistic models.

8 Conclusion

In conclusion, it is clear that based on the generated data in this study, there is no distinct answer as to the effectiveness of the bootstrapping methods. Throughout all of the analysis with one and two samples, the bootstrapping did not seem to perform significantly better in a lot of the cases. Additionally, contrary to the theory that Percentile-t is a better method than Percentile, the analysis showed some comparable variability with respect to this. Bootstrapping also did not show consistent findings when involved with censoring. At times it would perform better compared to no censoring and at other time not so.

It is however important to note some interesting conclusions. The bootstrapping seemed to perform better when the sample size was increased from $n = 20$ to $n = 100$ in the one sample results. Coverages went from $\approx 90\%$ to $\approx 94\%$ due to the increase in the sample size. $B = 100$ seemed to be a really small value and so it seemed to helped at times to have the other values of B. However, having having $B = 10,000$ was not really necessary, especially

since we continuously noticed that the differences between $B = 5,000$ and $B = 10,000$ was really small. Another surprising result was how the coverage got better with an increasing presence of censoring. Additionally, it was unexpected to see the Percentile and Percentile-t methods being so comparable, and how the Percentile-t was so different than Percentile and the Wald in the Channing House analysis.

In a future study, it would be worthwhile to incorporate results based on higher values of samples size to see whether bootstrapping would perform better and better as the size increases. Also, since it is said that bootstrapping has an advantage when it comes to robust model specification, it would be significant to perhaps perform this study without specifying a certain model.

9 Acknowledgments

I am very grateful for Dr. Susko, whom without his continued support and guidance this final paper would not be possible. With this being my last work in my Bachelor Degree in Statistics I would also like to express great gratitude to both my parents, Dr. Ammar Sarhan and Dr. Safaa Shahin, for their continuous support and encouragement throughout my degree.

10 References

John P. Klein and Melvin L, 2003, Moeschberger *Survival Analysis Techniques for censored and Truncated Data*.

A.C. Davison and D.V. Hinkley, 1997, *Bootstrap Methods an their Application*, Cambridge University Press.