P8160 Project 1 Final Report
A Simulation Study to Compare Three Survival Models

Jingchen Chai, Yi Huang, Ruihan Zhang

Contents

```
library(ggplot2)
```

Abstract

Survival analysis is a statistical method used to analyze time-to-event data, such as the time until a patient's death. It is commonly used in medical research to estimate the probability of an event occurring over time and to identify factors that may affect the risk of the event. The analysis also considers time-dependent covariates that vary over time and can be used to compare the survival times between different groups of patients. The objective of the simulation study is to evaluate the robustness of three survival models against the misspecified baseline hazard functions. To achieve this goal, we assess the accuracy and efficiency of the estimated treatment effects (β) for each model under various baseline hazard functions. This project will perform simulations to compare the parametric regression model (Exponential and Weibull) to the semi-parametric regression model (Cox) for survival data. We apply the inverse transformation method to generate survival data with censored observations from Exponential, Weibull, Gompertz, and Gamma distributions. Despite the different models, we found

Introduction

Survival analysis is a statistical technique used to examine data that measures the time it takes for an event to occur, such as a patient's death. It is commonly used in medical research to determine the likelihood of an event happening over a specific time period and to identify factors that may impact the risk of the event. This method takes into account censoring, which happens when participants do not experience the event before the end of the study or the event occurs after the study period. For instance, a study may track breast cancer patients from diagnosis to death or the end of a five-year period. In survival analysis, uncensored data is recorded when the event occurs at the exact observed event time, and it is coded as 1 in the status indicator variable. Conversely, censored data occurs when patients are lost to follow-up or the event happens after the study period, and it is coded as 0 in the status indicator variable. Time-dependent covariates are also considered in the analysis, which can vary over time and can be used to compare the survival times of different participant groups, such as a treatment group and a control group.

Objective

The objective is to design a simulation study to compare and contrast the efficiency and accuracy of the estimated treatment effects under different baseline hazard functions and assess their robustness against misspecified baseline hazard functions. A practical and effective recommendation will be provided to general users to select a suitable model on the basis of the numerical investigations.

Methodology

Section 1

Section 2

Plot