STA 3180 Statistical Modelling: Survival Analysis

STA 3180 Statistical Modelling: Lecture Notes on Survival Analysis

Introduction

Survival analysis is a branch of statistics that deals with the analysis of time-to-event data. It is used to analyze the time it takes for an event to occur, such as death or disease onset. It is also used to study the effects of treatments and other factors on the time to an event. Survival analysis is used in many fields, including medicine, engineering, economics, and social sciences.

Key Concepts

Survival analysis is based on the concept of survival function. The survival function is the probability that an individual will survive beyond a given time. It is usually expressed as a cumulative distribution function. The survival function can be estimated from observed data using methods such as the Kaplan-Meier estimator or the Cox proportional hazards model.

The hazard rate is the instantaneous rate at which an event occurs. It is the probability that an individual will experience the event at a given time, given that they have not experienced it before. The hazard rate can be estimated from observed data using methods such as the Cox proportional hazards model.

Coding Examples

Kaplan-Meier Estimator

```
Start of Code
# Load the library
library(survival)
# Create a data frame
df <- data.frame(time = c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10),
                                 event = c(1, 0, 0, 1, 0, 0, 0, 1, 0, 1))
# Fit the Kaplan-Meier estimator
km <- survfit(Surv(time, event) ~ 1, data = df)
# Plot the Kaplan-Meier estimator
plot(km, xlab = "Time", ylab = "Survival Probability")
End of Code
```

Cox Proportional Hazards Model

```
Start of Code
# Load the library
library(survival)
```

```
# Create a data frame
df <- data.frame(time = c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10),
                                event = c(1, 0, 0, 1, 0, 0, 0, 1, 0, 1),
                                age = c(20, 25, 30, 35, 40, 45, 50, 55, 60,
                                65))
# Fit the Cox proportional hazards model
coxph <- coxph(Surv(time, event) ~ age, data = df)
# Print the summary
summary(coxph)
End of Code
```

Practice Multiple Choice Questions

Q1. What is the survival function?

A. The probability that an individual will survive beyond a given time.