

surv_test

2023-06-16

Survival Analysis Tutorial using R

Introduction

Survival analysis is a statistical method used to analyze time-to-event data, where the event of interest could be, for example, death, failure, or any other event that occurs over time. In this tutorial, we will cover the fundamental concepts of survival analysis and demonstrate how to perform survival analysis using R.

Prerequisites

Before we begin, make sure you have the necessary R packages installed. We will be using the survival package for survival analysis, survminer for visualization, and dplyr for data manipulation. If you haven't installed these packages, you can do so using the following code:

```
# install.packages("survival")  
# install.packages("survminer")  
# install.packages("dplyr")  
  
source("libraries.R")
```

Dataset Description

For this tutorial, we will use a hypothetical dataset called “survival_data.csv.” The dataset contains the following variables:

time: The survival time in months. status: The event status, where 0 indicates censored and 1 indicates an event. age: The age of the subject at the start of the study. sex: The gender of the subject (male or female). treatment: The treatment group the subject belongs to (control or experimental). Make sure to load the dataset into R before proceeding with the analysis.

Data Preparation and Exploration

Let's start by creating the dataset and exploring its structure.

```
# Generate a synthetic survival dataset
set.seed(123)

# Number of observations
n <- 200

# Generate survival times (months)
time <- rexp(n, rate = 0.2) + 5

# Generate event status (0 = censored, 1 = event occurred)
event <- rep(1, n)
event[sample(1:n, n * 0.2)] <- 0

# Generate covariates
age <- rnorm(n, mean = 50, sd = 10)
sex <- sample(c("Male", "Female"), n, replace = TRUE)
treatment <- sample(c("Control", "Experimental"), n, replace = TRUE)

# Create the dataset
data <- data.frame(time = time, status = event, age = age, sex = sex, treatment = treatment)

# Print the first few rows of the dataset
head(data)
```

```
##      time status   age    sex  treatment
## 1  9.217      1 37.24 Female Experimental
## 2  7.883      0 73.75 Female Experimental
## 3 11.645      1 39.06 Female Experimental
## 4  5.158      1 51.92   Male      Control
## 5  5.281      1 48.74   Male Experimental
## 6  6.583      1 36.12 Female Experimental
```

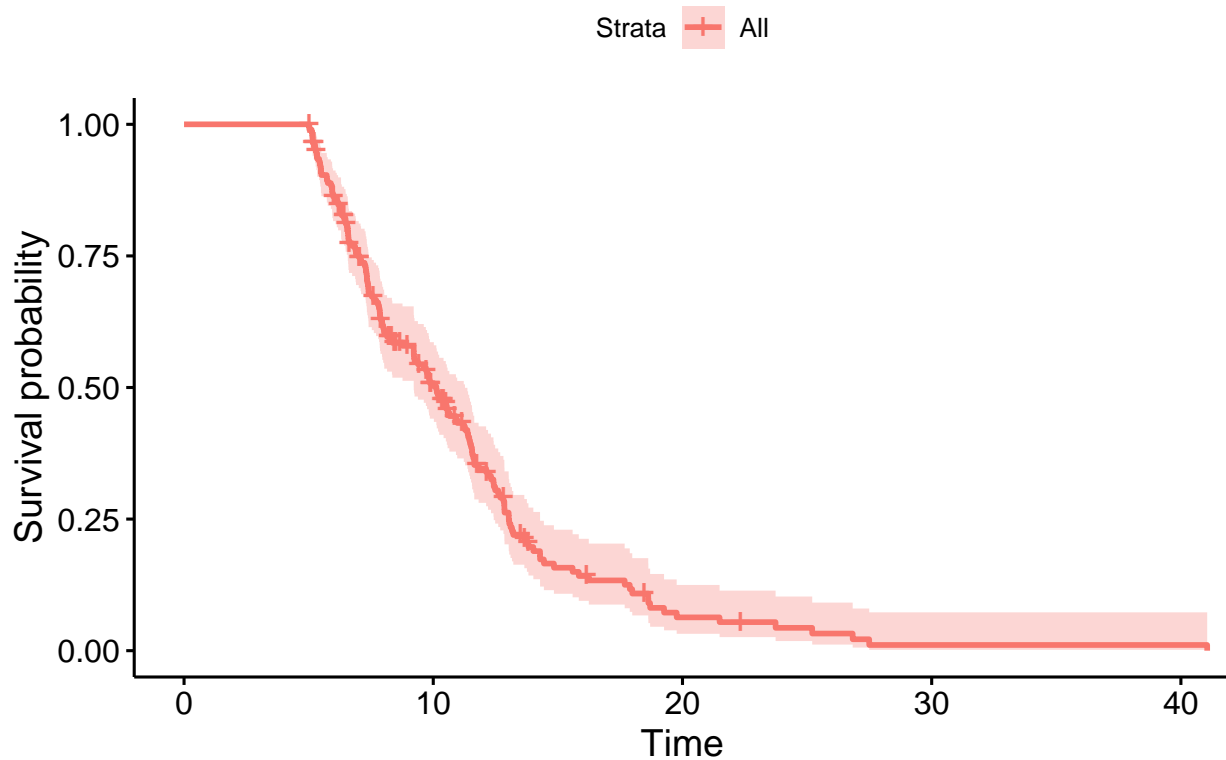
Kaplan-Meier Estimator and Survival Curves

The Kaplan-Meier estimator is a nonparametric method used to estimate the survival function from time-to-event data. We can plot survival curves using the `survfit()` and `ggsurvplot()` functions.

```
# Fit the Kaplan-Meier estimator
fit <- survfit(Surv(time, status) ~ 1, data = data)

# Plot the overall survival curve
ggsurvplot(fit, data = data, title = "Overall Survival Curve")
```

Overall Survival Curve



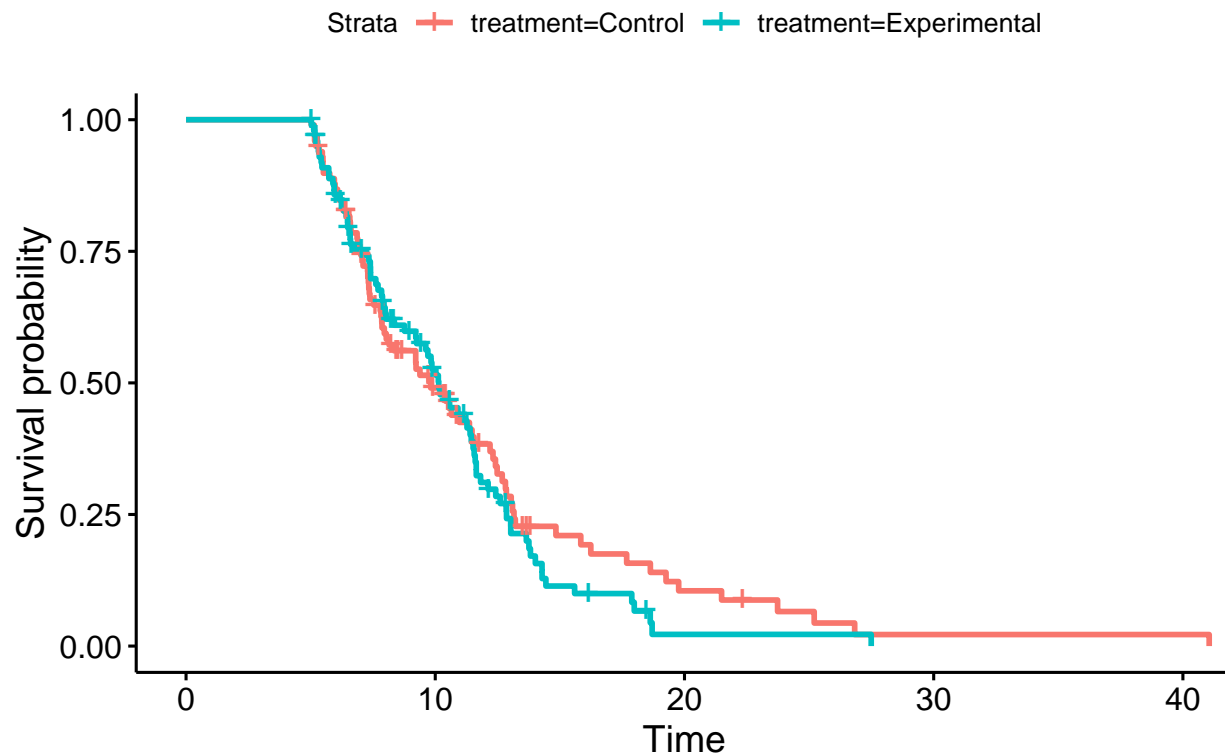
Stratified Survival Curves

We can also stratify the survival curves by different groups or covariates to analyze their impact on survival.

```
# Stratify the survival curves by treatment group
fit_stratified <- survfit(Surv(time, status) ~ treatment, data = data)

# Plot the stratified survival curves
ggsurvplot(fit_stratified, data = data, title = "Survival Curves by Treatment Group")
```

Survival Curves by Treatment Group



Log-Rank Test

The log-rank test is used to compare survival curves between different groups or categories. We can perform the log-rank test using the `survdif()` function.

```
# Perform the log-rank test
test <- survdiff(Surv(time, status) ~ treatment, data = data)

# Print the test results
summary(test)
```

```
##      Length Class  Mode
## n      2      table numeric
## obs    2      -none- numeric
## exp    2      -none- numeric
## var    4      -none- numeric
## chisq   1      -none- numeric
## pvalue  1      -none- numeric
## call   3      -none- call
```

Cox Proportional Hazards Model

The Cox proportional hazards model is a widely used regression model in survival analysis. We can fit the Cox model using the `coxph()` function.

```
# Fit the Cox proportional hazards model
cox_model <- coxph(Surv(time, status) ~ age + sex + treatment, data = data)

# Print the model summary
summary(cox_model)
```

```
## Call:
## coxph(formula = Surv(time, status) ~ age + sex + treatment, data = data)
##
##      n= 200, number of events= 160
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## age           -0.00743   0.99260  0.00800 -0.93   0.35
## sexMale        -0.03266   0.96787  0.16081 -0.20   0.84
## treatmentExperimental  0.13901   1.14913  0.16212  0.86   0.39
##
##              exp(coef) exp(-coef) lower .95 upper .95
## age              0.993         1.01    0.977    1.01
## sexMale           0.968         1.03    0.706    1.33
## treatmentExperimental  1.149         0.87    0.836    1.58
##
## Concordance= 0.513 (se = 0.026 )
## Likelihood ratio test= 1.64 on 3 df,  p=0.7
## Wald test              = 1.64 on 3 df,  p=0.6
## Score (logrank) test = 1.65 on 3 df,  p=0.6
```

Hazard Ratios and Significance

The Cox model provides hazard ratios, which measure the relative risk of an event for different covariate levels. We can interpret the hazard ratios and their significance using the model coefficients.

Conclusion

In this tutorial, we covered the fundamental concepts of survival analysis and demonstrated how to perform survival analysis using R. We explored the Kaplan-Meier estimator, stratified survival curves, log-rank test, and Cox proportional hazards model. Survival analysis is a powerful tool for analyzing time-to-event data and can provide valuable insights in various research fields.

Remember to adapt the code and analysis to your specific dataset and research question. Happy analyzing!