Survival Analysis

By: Yesilda Balavarca

International PhD Program
“Applied Statistics and Empirical Methods”
Introduction

Randomized Clinical Trial

Target population

Random assignment

Patient population

Treatment

Control

Event

Dead

Alive

Event

Dead

Alive

TIME
Introduction

- **Time to event**: Is the time from entry into a study until a subject has a particular outcome of interest (event)
  - Time to death
  - Time to relapse of a disease
  - Time to recovery from illness
  - Length of stay in a hospital
  - Time to finish a doctoral dissertation

- Kind of survival studies
  - Clinical trials
  - Prospective cohort studies
  - Retrospective cohort studies
Illustration of survival data
Time to event

- The time to event is always positive, non-negative → $T \geq 0$

- To correctly collect a time to event, we need:
  - an unambiguous time origin
  - a time scale
  - definition of the event

⇒ Is it always possible to record the time to event?
Censored observations

**Censoring**: Incomplete observation of the time to event.

- **Right censoring** \((T > t)\):
  - All we know is that during the period of observation there was no event
  - There is no time to event recorded because of:
    - loss to follow up:
      - drop out of the study
      - death due to a cause that is not the event of interest
    - termination of the study (the study ends before they have the outcome of interest)
Survival data - Example

- 10 patients with squamous cell carcinoma are recruited to receive specific treatment.
- The objective is to investigate the survival probability of the patients under this treatment.
  - Event of interest: death
- They are followed up to 16 months.
  - Duration of the study: 16 months
  - Time scale: months
- Consider right censored observations
Survival Data - unordered

Subject 1 — dies at 4 months
Subject 2 — drops out at 2 months
Subject 3 — dies at 6 months
Subject 4 — dies at 14 months
Subject 5 — dies at 1 month
Subject 6 — dies at 10 months
Subject 7 — had a mortal domestic accident at 5 months
Subject 8 — dies at 4 months
Subject 9 — survived for the whole study period
Subject 10 — dies at 3 months
Survival Data - ordered

<table>
<thead>
<tr>
<th>Subject</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>dies at 1 month</td>
</tr>
<tr>
<td>2</td>
<td>drops out at 2 months</td>
</tr>
<tr>
<td>10</td>
<td>dies at 3 months</td>
</tr>
<tr>
<td>1</td>
<td>dies at 4 months</td>
</tr>
<tr>
<td>8</td>
<td>dies at 4 months</td>
</tr>
<tr>
<td>7</td>
<td>had a mortal domestic accident at 5 months</td>
</tr>
<tr>
<td>3</td>
<td>dies at 6 months</td>
</tr>
<tr>
<td>6</td>
<td>dies at 10 months</td>
</tr>
<tr>
<td>4</td>
<td>dies at 14 months</td>
</tr>
<tr>
<td>9</td>
<td>survived for the whole study period</td>
</tr>
</tbody>
</table>

Outcome summary

<table>
<thead>
<tr>
<th>Censored</th>
<th>Event</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>7</td>
<td>10</td>
</tr>
</tbody>
</table>
Survival Analysis

- Survival analysis is concerned with studying the time between entry to a study and a subsequent event.
- Also called “time to event analysis”
- Survival analysis attempts to answer questions such as:
  - which fraction of a population will survive past a certain time?
  - at what rate will they fail?
    at what rate will they present the event?
  - How do particular factors benefit or affect the probability of survival?
Survival Analysis

Objectives

- To estimate time to event for a group of individuals.
- To compare time to event between two or more groups.
- To assess the relationship between explanatory variables and time to event.
Survival Analysis - advantages

1. Why not compare mean time to event between your groups using a t-test or linear regression?
   ⇒ ignores censoring

2. Why not compare proportion of events in your groups using logistic regression?
   ⇒ ignores time

Survival analysis accounts for censored observations as well as time to event.
Survival Analysis - methods

- Non-parametric estimation
  
  One group of patients
  - Kaplan-Meier
  - Life-Table (Actuarial Estimator)

  Comparison of groups
  - Logrank test
  - Linear rank logrank test

- Semi-parametric estimation model
  - Cox proportional hazard model (allows explanatory variables)

- Parametric models: exponential, weibull distribution, etc.
  (not covered here)
Kaplan - Meier survival method

- Non-parametric estimate of survival probability.
- Commonly used to describe survivorship of a study population.
- Intuitive graphical presentation.
- Cumulative survival characteristics.
- Estimation of median survival time.
- Commonly used to compare two study populations.
Kaplan - Meier Survival Curve

N° subjects at risk for death =10
Fraction surviving before 1 month =10/10 ⇒ 100%

Subject dies at 1 month
Hazard = 1/10
Fraction surviving =9/10

Subject drops out of the study at 2 months.
Subjects at risk for death = 8

Subject dies at 3 months
Hazard = 1/8
Fraction surviving =7/8
Kaplan - Meier Survival probability estimator

There are distinct event times \( t_1 < t_2 < \ldots < t_j \). At each time \( t_j \) there are \( n_j \) individuals still at risk and \( d_j \) individuals having the event.

Estimated survival probability at time \( t_j \):

\[
\hat{S}(t_j) = \prod_{j: t \leq t_j} \left[ 1 - \frac{d_j}{n_j} \right]
\]

Median Survival Time

Is the time \( t_j \) at which 50% of the patients have survived.

At the median survival time, half the subjects have developed the event and half are still on follow-up.

\[
\hat{S}(t_j) = 0.5
\]
Kaplan - Meier Survival probability

Survival probability at 3 months
= \frac{9}{10} \times \frac{7}{8} = 0.788

1-year survival rate \approx 28\%

Median survival time = 6 months
Comparison of groups – Logrank Test

Logrank-Test:

- For comparison of survival distributions between groups
- The groups are defined by categorical covariates.
  - e.g. Therapy: treatment, placebo
  - Gender: male, female
  - Age group: ≤40, ≥40

- Bad performance when two survival curves are crossing.
  - The logrank test has better performance under the assumption of proportional hazards.
Logrank Test - Example

Comparison of survival probability in patients with clear cell renal cell carcinoma: with and without lymph nodes involvement.

Log rank test:
p-value < 0.0001

The GNAS1 T393C Polymorphism Predicts Survival in Patients with Clear Cell Renal Cell Carcinoma.
Cox - Regression survival method

- Allows for **prognostic factors**.

- Explore the **relationship** between survival and explanatory variables.

- Models and compares the **hazards** for different groups/factors (explanatory variables).

- Important assumption:
  - Survival curves with **proportional hazards** (risk of an event at different time points).
Cox - Regression model

Also called: Cox Proportional Hazards model

\[
h_1(t, X) = h_0(t) \exp(\beta X)
\]

\[
h_1(t, X) = h_0(t) \exp(\beta_1 X_1 + \ldots + \beta_p X_p)
\]

t : time of event
X_i : set of covariates
\beta_i : set of parameters

Hazard ratio:

\[
\frac{h_1(t)}{h_0(t)} = \exp(\beta)
\]

- Constant, does not depend on time
- Proportional hazards over time

\exp(\beta) : Indicates how large (small) is the hazard in one group with the respect to the hazard in the reference group.
Cox - Regression model : Example

- HIV-patients receiving 2 types of treatment.

- The objective is to investigate the survival probability of the patients, by gender and by treatment.
  - Event of interest: death
  - Covariates: gender (male, female), treatment (new, standard)

- They are followed up to 6 years
  - Duration of the study: 6 years
  - Time scale: months

- Consider right censored observations
Cox - Regression model: exploring covariates

Gender

- Curves do not cross each other
- Proportional hazards …

Treatment

Good candidates to enter in the Cox-regression model
Cox - Regression model: parameter estimation

\[ h_1(t, X) = h_0(t) \exp(\beta_1 \text{gender} + \beta_2 \text{treatment}) \]

reference group: female, standard treatment

\[
\begin{align*}
\text{gender} &= \begin{cases} 
0 & \text{female} \\
1 & \text{male}
\end{cases} \\
\text{treatment} &= \begin{cases} 
0 & \text{standard} \\
1 & \text{new}
\end{cases}
\end{align*}
\]

\[ h_1(t, X) = h_0(t) \exp(-0.51 \text{gender} + 0.69 \text{treatment}) \]

\[ \exp(\beta_1) = \exp(-0.51) = 0.6 \quad \exp(\beta_2) = \exp(0.69) = 2.0 \]

Males have larger probabilities of survival than females. Patients receiving new treatment have lower survival probabilities than patients with standard treatment.
Thank you