BIST P8110: Applied Regression II

1. Introduction to Survival Data

Qixuan Chen

Department of Biostatistics
Columbia University

This lecture's big ideas

▶ Survival data
▶ Censoring
▶ Truncation

Introduction to survival data

- Commonly used names
    - survival data
    - censored data
    - time-to-event data
    - failure time data
- Numerous applications in biomedicines as well as engineering
    - time until HIV+ patient dies
    - time until cancer recurs
    - time until light-bulb goes out

Survival time

How to define survival time?

- ▶ The observation of survival time has 3 components
    - ▶ a begining point (i.e. when the "clock starts")
    - ▶ an endpoint (i.e. when the "clock stops")
    - ▶ the unit of time (days, weeks, months, years)
- ▶ **Survival time** is the distance on the time scale between the two points.

The event

What is the event?

- ▶ **Event** can mean death, disease, recurrence of disease, onset of symptoms etc.
- ▶ Event is commonly referred to as **failure**.

Example: Worcester Heart Attack Study (WHAS)

ID	Admission Date	Follow Up Date	Length of Stay	Follow up Time (days)	Vital Status
1	3/13/95	3/19/95	4	6	Dead
5	2/9/95	5/29/98	4	1205	Dead
10	7/22/95	12/31/02	9	2719	Alive
13	5/21/95	3/18/96	6	302	Dead

Table: Study of survival time following acute myocardial infarction among residents of Worcester, Massachusetts (*Hosmer et al. p.*4)

Censoring

- ▶ **Censored data:** Incomplete data, i.e., only partial information is known (a special type of missing data).
- ▶ **Right-censoring:** incomplete observation of a survival time due to
    - ▶ the end of the study or follow-up (e.g. the WHAS study, ID=10)
    - ▶ the patient is lost to follow up
    - ▶ the patient withdraws from the study
- ▶ Most survival data is **right-censored** and we will focus on methods for analyzing this type of data.

Other types of censoring

- **Left-censoring:** the event of interest has already occurred when observation begins
- **Interval-censoring:** data only known to lie between some thresholds.

Censoring: Example I

- ▶ **EX:** In a study to determine the distribution of the time until first marijuana use among high school students, the question was asked, "When did you first use marijuana?". A said "I have used it but can not recall just when the first time was". B and C said "I have never used it". One year later, B and C were asked the same questions again. This time, B said "Yes, I have used it" and C said "I still have never used it".

- ▶ **Question**: What is the event? How to define the survival time? Censoring or event for student A, B, and C? What type of censoring?

Censoring: Example II

- ▶ **EX:** In early childhood learning studies, interest often focuses upon testing children to determine when a child learns to accomplish certain specified tasks. Often, some children can already perform the task before the study starts. Some children may not learn the task during the entire study period.
- ▶ **Question**: What is the event? What is the survival time? What type of censoring for these two groups of children?

Truncation

- Truncation is another mechanism that leads to incomplete observation
    - Truncation of survival data occurs when only those individuals whose event time lies within a certain observational window (T_L, T_R) are observed.
    - An individual whose event time is not in this interval is not observed and no information on this subject is available to the investigators.
    - This is contrast to censoring where there is at least partial information on each subject.
- Because truncated-data occurs relatively infrequently in practice, we do not consider it in this class.

Left truncation

▶ **EX:** A common example of left truncation is the problem of estimating the distribution of the diameters of microscopic particles. Only particles that are big enough to be seen based on the resolution of a microscope are observed and smaller particles do not come to the attention of the investigator and are left truncated.

▶ As opposed to left censoring where we have partial information on individuals who experience the event of interest prior to T_L, for left truncation these individuals are never considered for inclusion into the study.

Right truncation

- ► **EX:** A common example of right truncation is the problem of estimating the distribution of the distances of stars from the earth. If a star is too far away to be visible, this star is right truncated.

- ► As opposed to right censoring where we have partial information on individuals who experience the event of interest post to T_R, for right truncation these individuals are never considered for inclusion into the study.

Truncation: Example

▶ **EX:** In a study of the waiting time from infection at transfusion to clinical onset of AIDS, the registry was sampled on June 30, 1986, so only patients whose waiting time from transfusion to AIDS was less than the time from transfusion to June 30, 1986, were available for observation. Patients transfused prior to June 30, 1986, who developed AIDS after June 30, 1986, were not observed.

▶ **Question:** What happened to patients who transfused prior to June 30, 1986 but developed AIDS after June 30, 1986?

Suggested Readings:

► Chapter 1 (Hosmer et al.)