Survival Analysis

2024.09.10

Textbooks

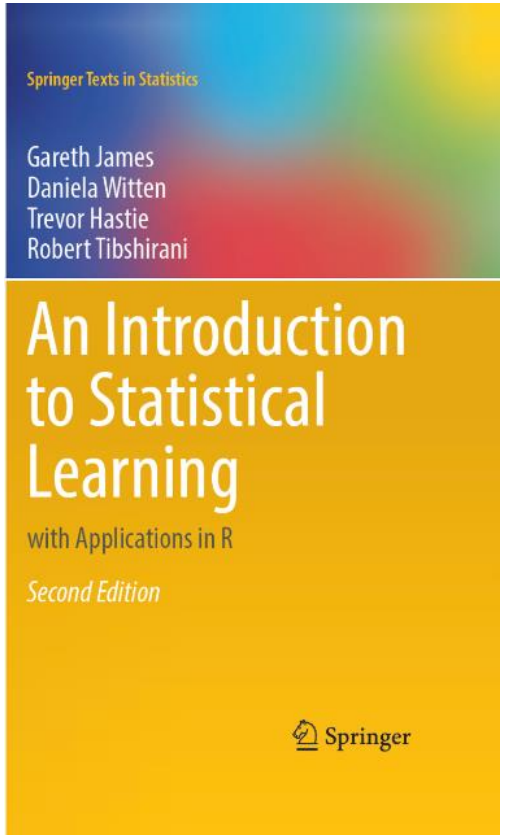
1. Klein, J.P. and Moeschberger, M.L. (2003) Survival analysis: Techniques for censored and truncated data. 2nd Edition, Springer, New York.

Data sets: R package “Kmsurv”

Textbooks (Lasso, Tree based method)

1. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani. (2013). An introduction to statistical learning : with applications in R. New York :Springer (Chapter 11)

https://www.statlearning.com/



Website for the Class materials

https://github.com/lhchien-ndhu/113-1-Survival-Analysis.git

Examples

Right censored data

1.11 Times to Death for Patients with Cancer of the Tongue

A study was conducted on the effects of ploidy on the prognosis of patients with cancers of the mouth. Patients were selected who had a paraffin-embedded sample of the cancerous tissue taken at the time of surgery. Follow-up survival data was obtained on each patient. The tissue samples were examined using a flow cytometer to determine if the tumor had an aneuploid (abnormal) or diploid (normal) DNA profile using a technique discussed in Sickle–Santanello et al. (1988). The data in Table 1.6 is on patients with cancer of the tongue. Times are in weeks.

The data is used in exercises.

tongue {KMsurv}

data from Section 1.11

Description

The `tongue` data frame has 80 rows and 3 columns.

Format

This data frame contains the following columns:

type

Tumor DNA profile (1=Aneuploid Tumor, 2=Diploid Tumor)

time

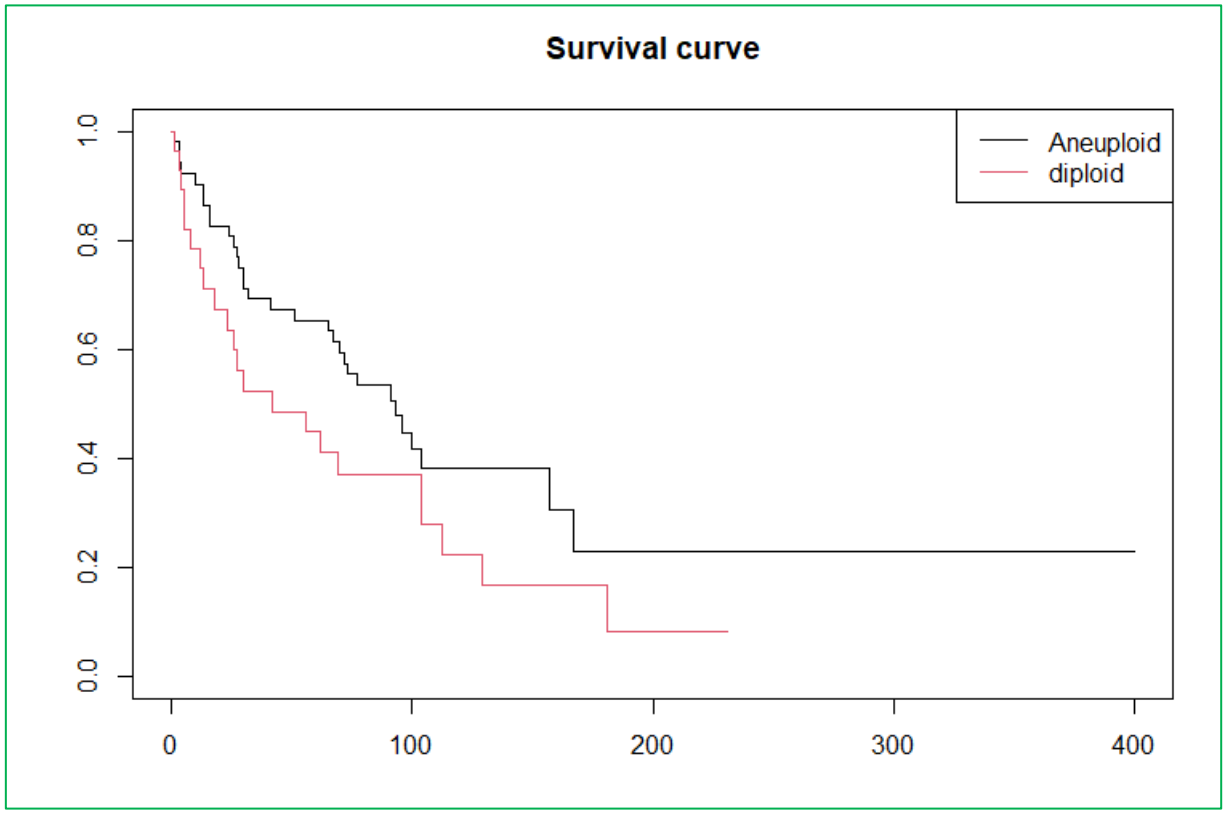
Time to death or on-study time, weeks

delta

Death indicator (0=alive, 1=dead)

```
> tongue
   type time delta
1     1    1     1
2     1    3     1
3     1    3     1
4     1    4     1
5     1   10     1
6     1   13     1
7     1   13     1
8     1   16     1
9     1   16     1
10    1   24     1
11    1   26     1
12    1   27     1
13    1   28     1
14    1   30     1
15    1   30     1
16    1   32     1
17    1   41     1
18    1   51     1
19    1   65     1
20    1   67     1
21    1   70     1
22    1   72     1
23    1   73     1
24    1   77     1
25    1   91     1
26    1   93     1
27    1   96     1
28    1  100     1
29    1  104     1
30    1  157     1
31    1  167     1
32    1   61     0
33    1   74     0
34    1   79     0
35    1   80     0
36    1   81     0
37    1   87     0
38    1   87     0
39    1   88     0
```

Right censored data



tongue {KMsurv}

data from Section 1.11

Description

The `tongue` data frame has 80 rows and 3 columns.

Format

This data frame contains the following columns:

type

Tumor DNA profile (1=Aneuploid Tumor, 2=Diploid Tumor)

time

Time to death or on-study time, weeks

delta

Death indicator (0=alive, 1=dead)

```
> tongue
   type time delta
1     1    1     1
2     1    3     1
3     1    3     1
4     1    4     1
5     1   10     1
6     1   13     1
7     1   13     1
8     1   16     1
9     1   16     1
10    1   24     1
11    1   26     1
12    1   27     1
13    1   28     1
14    1   30     1
15    1   30     1
16    1   32     1
17    1   41     1
18    1   51     1
19    1   65     1
20    1   67     1
21    1   70     1
22    1   72     1
23    1   73     1
24    1   77     1
25    1   91     1
26    1   93     1
27    1   96     1
28    1  100     1
29    1  104     1
30    1  157     1
31    1  167     1
32    1   61     0
33    1   74     0
34    1   79     0
35    1   80     0
36    1   81     0
37    1   87     0
38    1   87     0
39    1   88     0
```

Left truncation data

1.16 Death Times of Elderly Residents of a Retirement Community

Channing House is a retirement center located in Palo Alto, California. Data on ages at death of 462 individuals (97 males and 365 females) who were in residence during the period January 1964 to July 1975 has been reported by Hyde (1980). A distinctive feature of these individuals was that all were covered by a health care program provided by the center which allowed for easy access to medical care without any additional financial burden to the resident. The age in months when members of the community died or left the center and the age when individuals entered the community is available on the authors' web site.

The life lengths in this data set are *left truncated* because an individual must survive to a sufficient age to enter the retirement community. Individuals who die at an early age are excluded from the study. Ignoring this left truncation leads to the problem of length-biased sampling. The concept of left truncation and the bias induced into the estimation process by ignoring it is discussed in section 3.4.

This data will be used in section 4.6 to illustrate how one estimates the conditional survival function for left-truncated data. The data is used in section 7.3 to illustrate the comparison of two samples (male and female), when there is left truncation and right censoring employing the log-rank test, and in Chapter 9 employing the Cox proportional hazards model.

channing {KMsurv}

data from Section 1.16

Description

The `channing` data frame has 462 rows and 6 columns.

Format

This data frame contains the following columns:

obs
: Observation number

death
: Death status (1=dead, 0=alive)

ageentry
: Age of entry into retirement home, months

age
: Age of death or left retirement home, months

time
: Difference between the above two ages, months

gender
: Gender (1=male, 2=female)

```
> head(channing,10)
   obs death ageentry  age time gender
1    1     1     1042 1172  130      2
2    2     1      921 1040  119      2
3    3     1      885 1003  118      2
4    4     1      901 1018  117      2
5    5     1      808  932  124      2
6    6     1      915 1004   89      2
7    7     1      901 1023  122      2
8    8     1      852  908   56      2
9    9     1      828  868   40      2
10  10     1      968  990   22      2
```

Competing event

- 1.13 Bone Marrow Transplantation for Leukemia

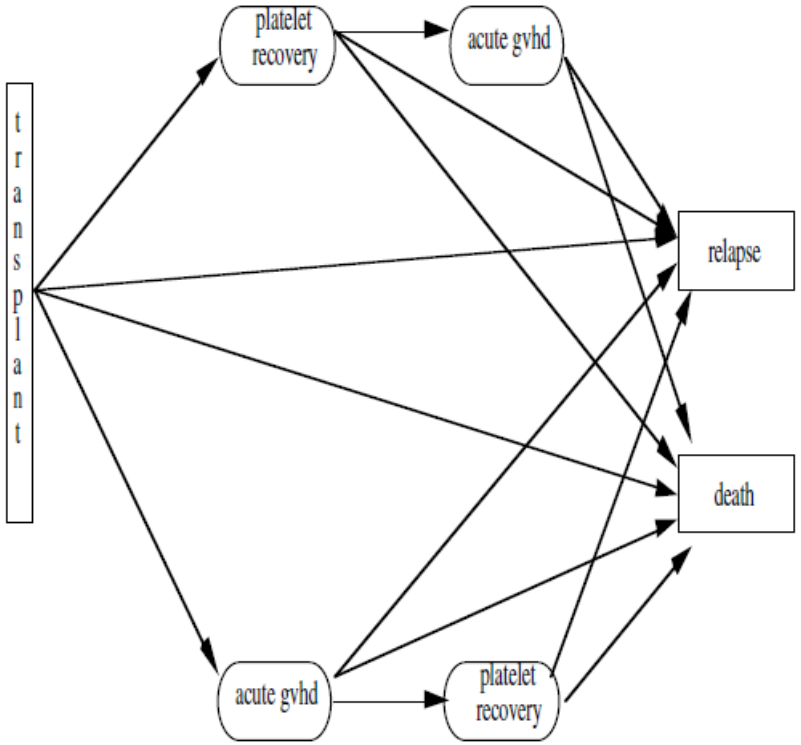


Figure 1.1 *Recovery Process from a Bone Marrow Transplant*