

Lin_ST625_HW1

Frances Lin

1/5/2022

1a

Time to leukemia relapse is the variable of interest.

1b

Some events may not be observed within the study time period. More specifically, censoring is needed when, within a given study time frame, we only have partial information about the survival time of some individuals, but we don't know their true survival time.

For example, we need to censor patients who do not show relapse at the end of the study since we only know that these patients do not relapse, but we don't know the exact relapse time of these patients, if they ever relapse. We also need to censor those who exit the study before the events are observed.

1c

A paired t-test would do since we are comparing the mean survival times of two groups and the patients are in pairs.

1d

The quality of results can be undermined (e.g. Distribution may be skewed). Even though traditional models such as that mentioned in 1c can be used to study the occurrence of an event, they do not account for the timing of the occurrence.

2a

Group A

2b

Group B

2c

Group A ($\hat{S}(t) = 0.5$ is $t = 5$ for group A and $t = 3$ for group B.)

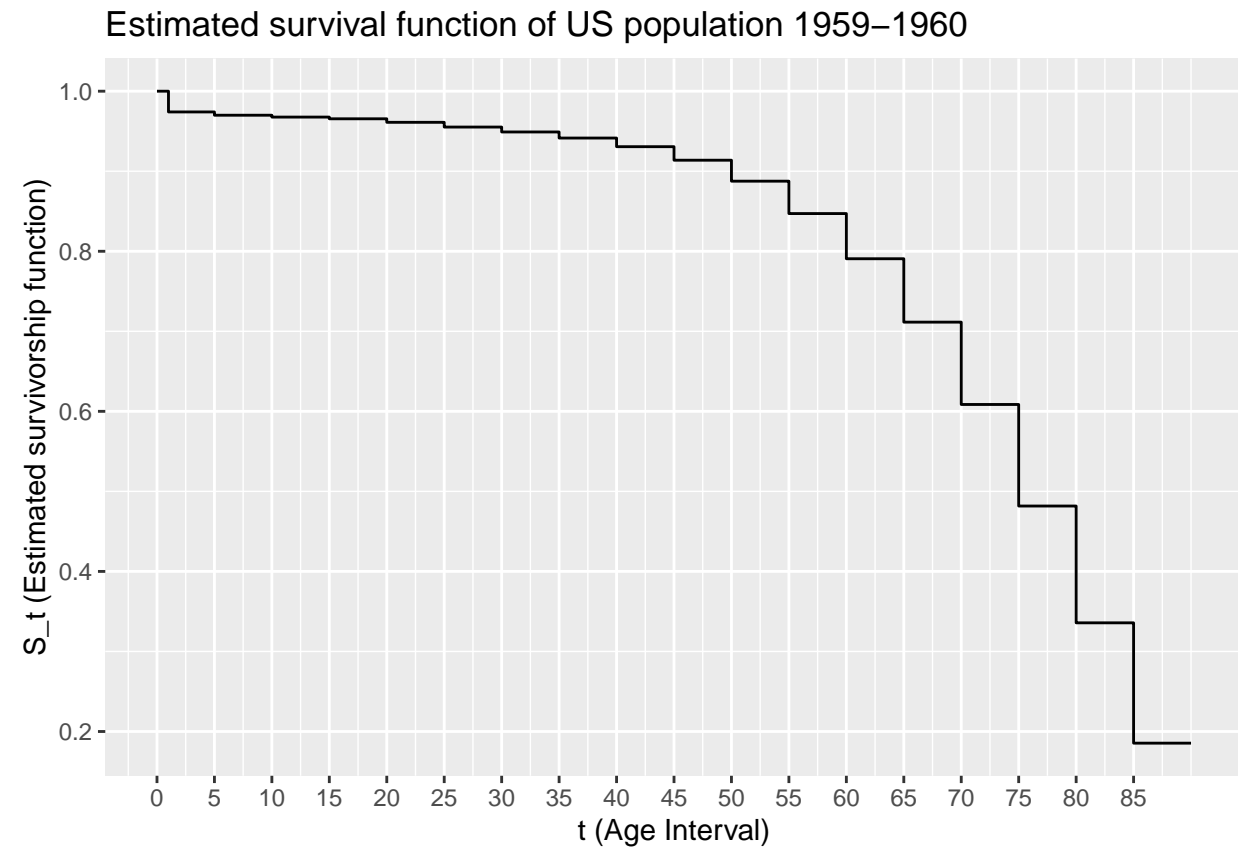
3a

S_t = #s of people surviving longer than t / total # of people and

f_t = #s of people dying in the interval beginning at time t / ((total # of patients) x (interval width)).

For example, $S(t = 1) = \frac{100000}{100000} = 1$, $S(t = 2) = \frac{97407}{100000} = 0.97407$, etc. $f(t = 1) = \frac{2593}{100000 \times (2-1)} = 0.02593$, etc.

Age	Age_2	Living	Dying	S_t	f_t
0	1	1e+05	2593	1	0.02593
1	5	97407	409	0.9741	0.001022
5	10	96998	233	0.97	0.000466
10	15	96765	214	0.9677	0.000428
15	20	96551	440	0.9655	0.00088
20	25	96111	594	0.9611	0.001188



3b

$S(t = 70) = 0.60857$.

Age	Age_2	Living	Dying	S_t	f_t
70	75	60857	12687	0.6086	0.02537

4a

age	sex	death	time	status
21.9	2	0	424	1
24	1	0	1241	0
29.6	1	0	280	0
22.8	2	0	570	1
29.7	2	0	65	0
23.2	2	0	1344	0

4b

age	time
Min. :14.60	Min. : 0.0
1st Qu.:22.30	1st Qu.: 434.0
Median :24.30	Median : 767.0
Mean :24.25	Mean : 901.8
3rd Qu.:26.30	3rd Qu.:1189.0
Max. :30.50	Max. :5070.0

51.99% of women whose first child is a boy (sex = 1).

sex	n	percent
1	27842	51.99
2	25715	48.01

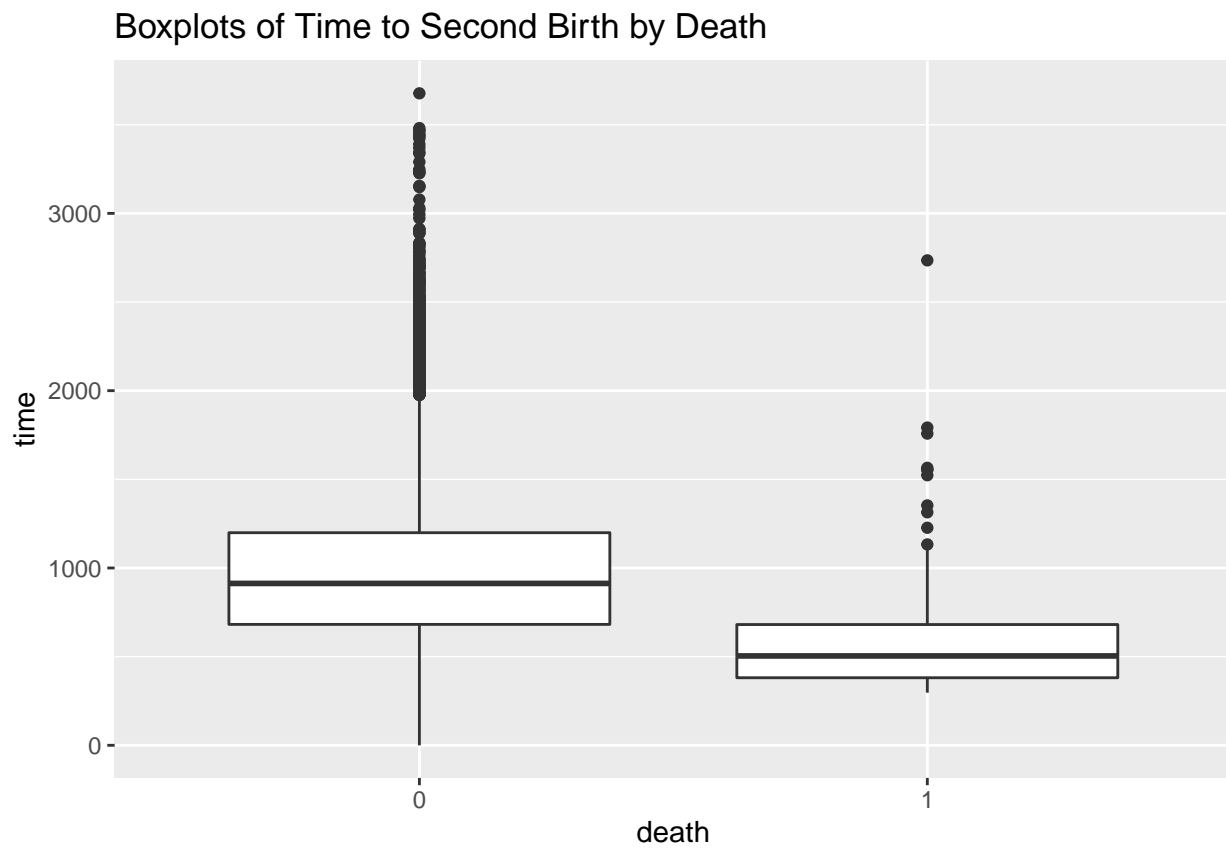
0.4892% of women whose first child died within one year (death = 1).

death	n	percent
0	53295	99.51
1	262	0.4892

69.49% of data was censored (status = 0).

status	n	percent
0	37216	69.49
1	16341	30.51

4c



4d

We cannot infer because we are not fitting a model, although for women with second birth observed, if the first child dies within one year of its birth, time from first birth to second birth or censoring appears to be shorter.