

Common used function and commands in R

- Survival Analysis

- let $f(t)$ be the probability density function (p.d.f) of event time T . Then $F(t)$ represents $S(t) = \Pr(T > t) = 1 - \Pr(T \leq t) = 1 - \int_0^t f(u) du$

$$a^2 + b^3 + \frac{1}{2}$$

- There are three main types of censoring: **right, left, and interval**. Censoring could occur, for example, when administering a survey to mothers every other month asking if they are still breast feeding. Right censoring occurs when mothers are still breast feeding after the last survey, since we do not know exactly how long they will continue. Left censoring occurs when mothers enter the study after they have stopped breast feeding. We do not know exactly when they stopped breast feeding, although we know that it happened before their entry to the study. Interval censoring occurs if the breast feeding ended between two successive surveys since one can only say that breast feeding ended somewhere within the past two months.
- Simple approaches researchers might choose to deal with censored data are to set the censored observations to missing or replace the unobserved value of the variable by zero, the minimum, maximum, mean value, or a randomly assigned value from the range of possible values. When the censoring is minimal, using one of these approaches can be reasonable. When it is not, these simple solutions can, however, cause serious bias in estimates and standard errors obtained in subsequent statistical analysis, can discard potent
- The logistic regression, for example, can be applied to estimate the probability of experiencing a particular lifetime event within a limited time period; nevertheless, it does not consider the time when the event occurs and therefore disregards the length of the survival process. Suppose that two population groups have the same rate of experiencing a particular event by the end of an observation period but members in one group are expected to experience the event significantly later than do those in the other. The former population group has an advantaged survival pattern because its average life is extended. Obviously, the logistic regression ignores this timing factor, therefore not providing precise information.
- Kaplan-Meier survival estimates

Table 2.1 Kaplan–Meier survival estimates for older women with breast cancer.

Month	Calculating steps	Survival rate
0	$10/10 = 1.0$	1.0
5	$1.0 \times (9/10) = 0.9$	0.9
17	$1.0 \times (9/10) \times (8/9) = 0.8$	0.8
24	$1.0 \times (9/10) \times (8/9) \times (7/8) = 0.7$	0.7
32	$1.0 \times (9/10) \times (8/9) \times (7/8) \times (6/7) = 0.6$	0.6
40	$1.0 \times (9/10) \times (8/9) \times (7/8) \times (6/7) \times (5/6) = 0.5$	0.5
46	$1.0 \times (9/10) \times (8/9) \times (7/8) \times (6/7) \times (5/6) \times (4/5) = 0.4$	0.4
47	$1.0 \times (9/10) \times (8/9) \times (7/8) \times (6/7) \times (5/6) \times (4/5) \times (3/4) = 0.3$	0.3

Table 2.2 Kaplan–Meier survival estimates with right censoring.

Month	Calculating steps	Survival rate
0	$12/12 = 1.0$	1.00
5	$1.0 \times (11/12) = 0.92$	0.92
17	$1.0 \times (11/12) \times (10/11) = 0.83$	0.83
20*	$1.0 \times (11/12) \times (10/11) \times (9/9) = 0.83$	0.83
24	$1.0 \times (11/12) \times (10/11) \times (9/9) \times (8/9) = 0.74$	0.74
32	$1.0 \times (11/12) \times (10/11) \times (9/9) \times (8/9) \times (7/8) = 0.65$	0.65
35*	$0.65 \times (6/6) = 0.65$	0.65
40	$0.65 \times (6/6) \times (5/6) = 0.54$	0.54
46	$0.65 \times (6/6) \times (5/6) \times (4/5) = 0.43$	0.43
47	$0.65 \times (6/6) \times (5/6) \times (4/5) \times (3/4) = 0.32$	0.32