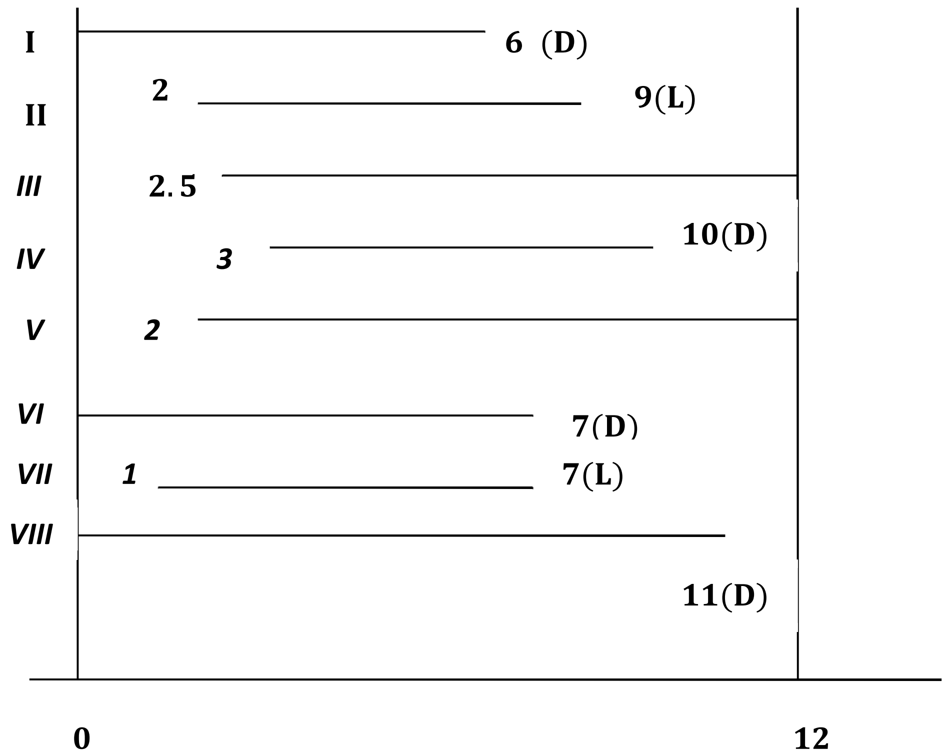
Chapter 5 Models for Survival Data

5.1 Introduction:

1. Motivation

Motivating example:

There are 8 patients (I~VIII) in a 12-months clinical study (D: death; L: loss information).



In this data, the exact failure times of some patients are unknown either because the patients withdraw from the study or because the patients were still alive at the end of the study. We refer the above situation as "censoring". Censoring is so common in medical experiments that the statistical methods must allow for it.

For the above example, we have the following data:

6	7^*	9.5^*	7	10^*	7	6^*	11

where "*" stands for censoring. Further, if we have the following information about these patients:

Subject	Survival time	Censor indicator	Group	# of cigarette	Gender	Age
I	6	1	T	20	1	45
II	7	0	T	30	1	20
III	9.5	0	T	5	0	38
IV	7	1	T	40	1	26
V	10	0	C	3	0	42
VI	7	1	C	40	0	17
VII	6	0	C	60	1	25
VIII	11	1	C	10	0	29

where Censor indicator (1: death; 0: censoring) and Group (T: treatment group; C: control group).

Objective:

For the above typical survival data, we are concerned with

1. the survival function
2. the comparison of two groups of survival data
3. which factors (# of cigarette, gender or age) are important in deciding the failure rate

2. Survival functions and hazard functions

Let T be the positive random variable representing failure time. Then,

$$S(t) = P(T \geq t) = \int_t^{\infty} f(x)dx$$

is called the survival function, where $f(x)$ is the probability density function of T.

The hazard function is

$$\begin{aligned}
\lambda(t) &= \lim_{\varepsilon \to 0^+} \left[\frac{P(t \leq T < t + \varepsilon | T \geq t)}{\varepsilon} \right] \\
&= \lim_{\varepsilon \to 0^+} \left[\frac{1}{\varepsilon} \cdot \frac{P(t \leq T < t + \varepsilon)}{P(T \geq t)} \right] \\
&= \lim_{\varepsilon \to 0^+} \left[\frac{1}{\varepsilon} \cdot \frac{\int_t^{t+\varepsilon} f(x)dx}{S(t)} \right] \\
&= \frac{1}{S(t)} \cdot \lim_{\varepsilon \to 0^+} \left[\frac{\int_t^{t+\varepsilon} f(x)dx}{\varepsilon} \right] \\
&= \frac{f(t)}{S(t)}
\end{aligned}$$

The relationships among $\lambda(t), f(t), S(t)$:

1.

$$\lambda(t) = -\frac{dlog[S(t)]}{dt}.$$

2.

$$S(t) = exp\left[-\int_0^t \lambda(x)dx\right].$$

3.

$$f(t) = \lambda(t)S(t) = \lambda(t)exp\left[-\int_0^t \lambda(x)dx\right].$$

[derivation:]

1.

$$-\frac{dlog[S(t)]}{dt} = -\frac{1}{S(t)} \cdot \frac{dS(t)}{dt} = -\frac{1}{S(t)} \cdot \frac{d[\int_t^\infty f(x)dx]}{dt}$$

$$= -\frac{1}{S(t)} \cdot \frac{d\left[1 - \int_0^t f(x)dx\right]}{dt} = -\frac{1}{S(t)} \cdot (-1) \cdot f(t) = \frac{f(t)}{S(t)}$$

$$= \lambda(t).$$

2. By 1,

$$\lambda(t) = -\frac{dlog[S(t)]}{dt}$$

$$\begin{aligned} \Leftrightarrow -\int_0^t \lambda(x)dx &= \int_0^t \left\{\frac{dlog[S(x)]}{dx}\right\} dx = \int_0^t dlog[S(x)] \\ &= log[S(x)]|_0^t = log[S(t)] - log[S(0)] \\ &= log[S(t)] - log[P(T \geq 0)] \\ &= log[S(t)] - log(1) \\ &= log[S(t)] \end{aligned}$$

$$\Leftrightarrow S(t) = exp\left[-\int_0^t \lambda(x)dx\right]$$

3. By 2, since

$$\lambda(t) = \frac{f(t)}{S(t)},$$

then

$$f(t) = \lambda(t)S(t) = \lambda(t) \cdot exp\left[-\int_0^t \lambda(x)dx\right].$$

Example:

(i) Constant hazards: $\lambda(t) = \lambda$

Then,

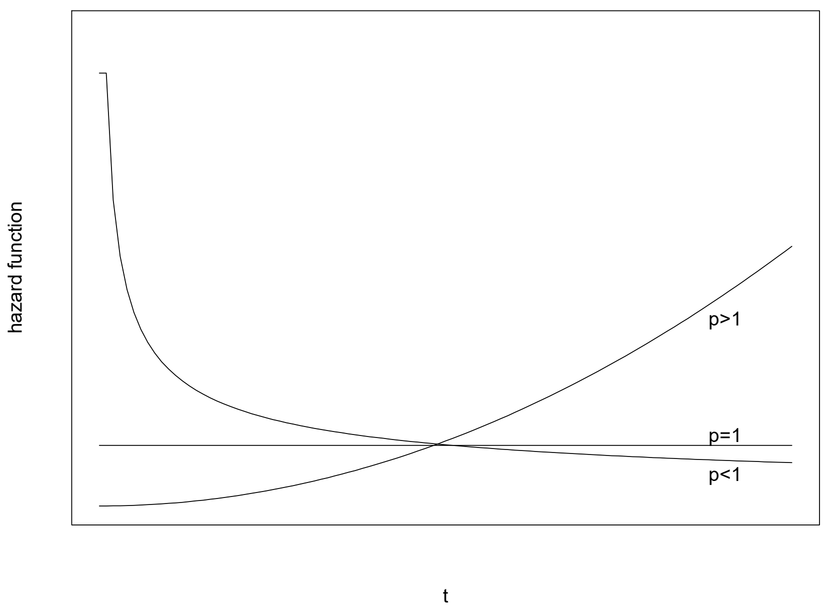
$$S(t) = exp\left[-\int_0^t \lambda(x)dx\right] = exp\left(-\int_0^t \lambda dx\right) = exp(-\lambda t)$$

and

$$f(t) = \lambda(t) \cdot exp\left[-\int_0^t \lambda(x)dx\right] = \lambda \cdot exp\left(-\int_0^t \lambda dx\right) = \lambda \cdot exp(-\lambda t)$$

is the exponential distribution.

(ii) $\lambda(t) = \lambda p(\lambda t)^{p-1}$



Then,

$$\boldsymbol{S(t) = exp\left[-\int_0^t \lambda(x)dx\right] = exp\left[-\int_0^t \lambda p(\lambda x)^{p-1}dx\right]}$$

$$\boldsymbol{= exp\left[-\int_0^t p(\lambda x)^{p-1}d\lambda x\right] = exp[-(\lambda t)^p]}$$

and

$$\boldsymbol{f(t) = \lambda(t)\cdot exp\left[-\int_0^t \lambda(x)dx\right] = \lambda p(\lambda t)^{p-1}\cdot exp[-(\lambda t)^p]}$$

is the Weibull distribution.

(iii)

$$\boldsymbol{\lambda(t) = \alpha t + \frac{\beta}{1+rt}}$$

is bathtub hazard function.