BHARATHIDASAN UNIVERSITY

Tiruchirappalli- 620024

Tamil Nadu, India.

Programme: M.Sc. Statistics

Course Title: Survival Analysis and Clinical Trials

Course Code: 23ST04DEC

Unit-I

Introduction to Survival Analysis

Dr. T. Jai Sankar
Associate Professor and Head
Department of Statistics

Ms. I. Angel Agnes Mary
Guest Faculty
Department of Statistics

UNIT – I

Survival Analysis

Introduction to Survival Analysis

Survival time can be defined broadly as the time to the occurrence of a given event. The term 'survival time' specifies the length of time taken for failure to occur.

This **event** can be the development of a disease, response to a treatment, relapse, or death. Therefore, survival time can be tumor-free time, the time from the start of treatment to response, length of remission, and time to death. Survival data can include survival time, response to a given treatment, and patient characteristics related to response, survival, and the development of a disease.

Survival analysis is the name for a collection of statistical procedures for data analysis for which the outcome variable of interest is time until an event occurs. In survival analysis we use the term 'failure' to define the occurrence of the event of interest (even though the event may actually be a 'success' such as recovery from therapy). Situations where survival analyses have been used in epidemiology include:

- Survival of patients after surgery.
- The length of time taken for cows to conceive after calving.
- The time taken for a farm to experience its first case of an exotic disease.

Example is study that follows patients who received a heart transplant to find out how long these patients survive after receiving the transplant. Another example is a study that follows a group of persons who are initially disease-free over several years to see who develops heart disease. To analyze data from these types of studies, a collection of statistical procedures called **survival analysis** is needed.

The above examples are survival analysis problems because the outcome variable is time until an event occurs. In the first example, the event is "developing heart disease," and the outcome is "time in years until a person develops heart disease." In the Second example, the event is "death" and the outcome is "time in years until death."

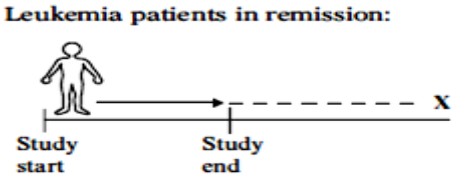
Censored data

Most survival analyses must consider a key analytical problem called censoring. In essence, censoring occurs when we have some information about individual survival time, but we don't know the survival time exactly.

As a simple example of censoring, consider leukemia patients followed until they go out of remission, shown here as X. If for a given patient, the study ends while the patient is still in remission (i.e., doesn't get the event), then that patient's survival time is considered censored. We know that, for this person, the survival time is at least as long as the period that the person has been followed, but if the person goes out of remission after the study ends, we do not know the complete survival time.

Remission refers to a temporary period during which an illness becomes less censor.

Relapse refers to fall back into worst state after improvement.

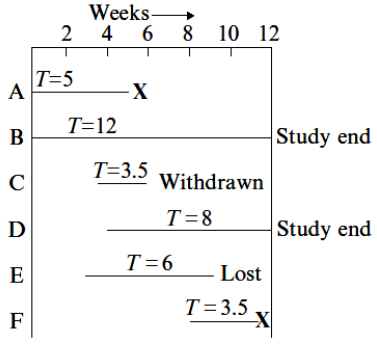


There are generally three reasons why censoring may occur:

- a person does not experience the event before the study ends;
- a person is lost to follow-up during the study period;
- a person withdraws from the study because of death (if death is not the event of interest) or some other reason (e.g., adverse drug reaction or other competing risk)

These situations are graphically illustrated here. The graph describes the experience of several persons followed over time. An X denotes a person who got the event.

- Person A is followed from the start of the study until getting the event at week 5; his survival time is 5 weeks and is not censored.
- Person B also is observed from the start of the study but is followed to the end of the 12-weekstudy period without getting the event; the survival time here is censored because we can say only that it is at least 12 weeks.
- Person C enters the study between the second and 3rd week and is followed until he withdraws from the study at 6 weeks; this person's survival time is censored after 3.5 weeks.
- Person D enters at week 4 and is followed for the remainder of the study without getting the event; this person's censored time is 8 weeks.
- Person E enters the study at week 3 and is followed until week 9, when he is lost to follow-up; his censored time is 6 weeks.
- Person F enters at week 8 and is followed until getting the event at week 11.5. As with person A, there is no censoring here; the survival time is 3.5 weeks.



In summary, of the six persons observed, two get the event (persons A and F) and four are censored (B, C, D, and E).

Event: A, F

Censored: B, C, D, E

A table of the survival time data for the six persons in the graph is now presented. For each person, we have given the corresponding survival time up to the event's occurrence or upto censorship. We have indicated in the last column whether this time was censored or not (with 1 denoting failed and 0 denoting censored).

For example, the data for person C is a survival time of 3.5 and a censorship indicator of0, whereas for person F the survival time is 3.5and the censorship indicator is 1. This table is a simplified illustration of the type of data to be analyzed in a survival analysis.

Person	Survival time	Failed (1); Censored (0)
A	5	1
B	12	0
C	3.5	0
D	8	0
E	6	0
F	3.5	1

Types of censored

There are three types of censored.

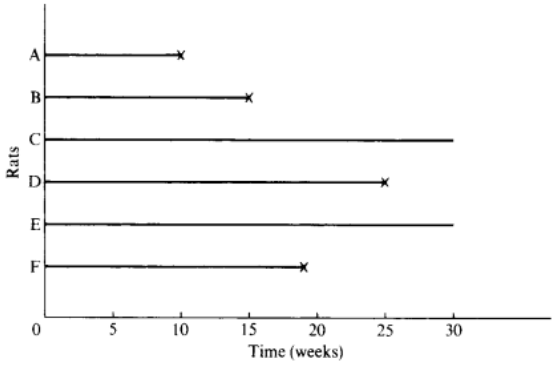
- Type I censoring
- Type II censoring
- Type III censoring

Type I censoring

Studies based on animal usually start with a fixed number of animals, to which the treatment or treatments is given. Because of time and cost limitations, the researcher often cannot wait for the death of all the animals. One option is to observe for a fixed period of time, say six months. Survival times recorded for the animals that died during the study period are the times from the start of the experiment to their death. These are called exact or uncensored observations. The survival times of the remaining animals are not exactly known but they are recorded as at least the length of the study period. These are called censored observations. Some animals could be lost or die accidentally. Their survival times, from the start of experiment to loss or death, are also censored observations. In type I censoring, if there are no accidental losses, all censored observations equal the length of the study period.

For example, suppose that six rats have been exposed to carcinogens by injecting tumor cells into their foot pads. The times to develop a tumor of a given size are observed. The investigator decides to terminate the experiment after 30 weeks. Rats A, B, and D developed

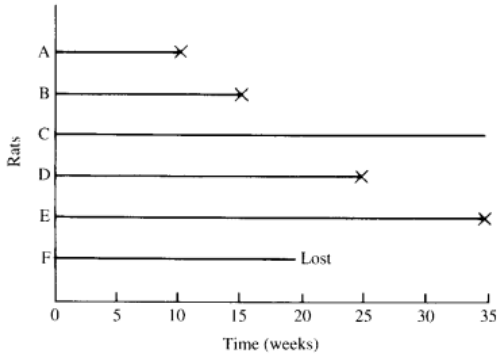
tumors after 10, 15, and 25 weeks, respectively. Rats C and E did not develop tumors by the end of the study. Rat F died accidentally without any tumors after 19 weeks of observation. The survival data are 10, 15, 30+, 25, 30+and 19+ weeks. (The plus sign indicates the censored observation.)



Example of type I censored data.

Type II censoring

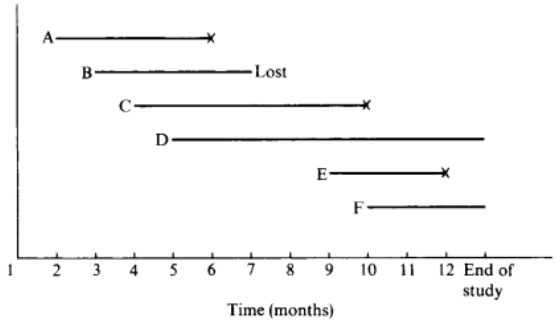
Another option in animal studies is to wait until a fixed portion of the animals have died, say 80 of 100, after which the study is terminated. In type II censoring, if there are no accidental losses, the censored observations equal the largest uncensored observation. For example, in the previous example, investigator may decide to terminate the study after four of the six rats have developed tumors. The survival or tumor-free times are then 10, 15, 35+, 25, 35, and 19+ weeks.



Example of type II censored data.

Type III censoring

In most clinical studies the period of study is fixed and patients enter the study at different times during that period. Some patients may die before the end of the study; their exact survival times are known. Others may withdraw before the end of the study and are lost to follow-up. Some patients might still be alive at the end of the study. For example, suppose that six patients with acute leukemia enter a clinical study during a total study period of one year. Patients A, C and E achieve remission at the beginning of the second, fourth, and ninth months and relapse after four, six and three months, respectively. Patient B achieves remission at the beginning of the third month but he is lost to follow-up four months later. Patients D and F achieve remission at the beginning of the fifth and tenth months, respectively, and they are still in remission at the end of the study. Hence the respective remission times of the six patients are 4, 4+, 6, 8+, 3 and 3+ months.



Example of type III censored data.

Type I and type II censored observations are also called singly censored data, and type III, progressively censored data. Another commonly used name for type III censoring is random censoring. All of these types of censoring are right censoring or censoring to the right. There are also left censoring and interval censoring cases. Left censoring occurs when it is known that the event of interest occurred prior to a certain time t, but the exact time of occurrence is unknown.

For example, an epidemiologist wishes to know the age at diagnosis in a follow-up study of diabetic retinopathy. At the time of the examination, a 50-year-old participant was found to have already developed retinopathy, but there is no record of the exact time at which initial evidence was found. Thus the age at examination (i.e., 50) is a **left-censored observation**. It means that the age of diagnosis for this patient is at most 50 years. Interval censoring occurs when the event of interest is known to have occurred between times a and b. For example, if medical records indicate that at age 45, the patient in the example above did not have retinopathy; his age at diagnosis is between 45 and 50 years.

Likelihood

The likelihood is the probability or probability density of what was observed, viewed as a function of parameters in an assumed model. To incorporate censored data points in the likelihood the censored data points are represented by the probability of the censored data points as a function of the model parameters given a model, i.e. a function of CDF(s) instead of the density or probability mass.

The most general censoring case is interval censoring

$$\Pr(a < x \leq b) = F(b) - F(a)$$

Where, *F(x)* is the CDF of the probability distribution, and the two special cases are:

- left censoring:

$$\Pr(-\infty < x \leq b) = F(b) - F(-\infty) = F(b) - 0 = F(b) = \Pr(x \leq b)$$

- right censoring:

$$\Pr(a < x \leq \infty) = F(\infty) - F(a) = 1 - F(a) = 1 - \Pr(x \leq a) = \Pr(x > a)$$

For continuous probability distributions

$$\Pr(a < x \leq b) = \Pr(a < x < b)$$

Likelihood function under censoring and estimation

The likelihood under censoring can be constructed using either the density and distribution functions or the hazard and cumulative hazard functions. Both are equivalent. The log likelihood will be a mixture of probabilities and densities, depending on whether the observation was censored or not.

Suppose we are interested in survival times, $T_1, T_2, \ldots, T_n$, but we don't observe T_i for all i. Instead, we observe

(U_i, δ_i), with $U_i = T_i$ and $\delta_i = 1$ if T_i is actually observed and

(U_i, δ_i), with $U_i < T_i$ and $\delta_i = 0$ if all we know is that T_i is longer than U_i.

when $T_i > U_i$, U_i is called the censoring time.

If the censoring times are all known constants, then the likelihood is

$$L = \prod_{i,,\delta_i=1} f(u_i) \prod_{i,,\delta_i=0} S(u_i)$$

where $f(u_i)$ = the probability density function evaluated at u_i

and $S(u_i)$ = the probability that T_i is greater than u_i, called the survival function.

This can be simplified by defining the hazard function, the instantaneous force of mortality, as

$$\lambda(u) = f(u) \Big/ S(u)$$

$$f(u) = \lambda(u)\, S(u).$$

Then

$$L = \prod_i \lambda(u_i)^{\delta_i}\, S(u_i)$$

For the exponential distribution, this becomes even simpler, because the hazard rate, λ, is constant, and$S(u) = exp(-\lambda u)$. Then:

$$L(\lambda) = \lambda^k \exp(-\lambda \sum u_i)$$

where $k = \sum \delta_i$.

From this we compute $\hat{\lambda}$, the maximum likelihood estimate (MLE) of λ, as follows:

$$l(\lambda) = \log(L(\lambda)) = k \log(\lambda) - \lambda \sum u_i$$

Then

$$dl/d\lambda = k/\lambda - \sum u_i$$

We set this to 0 and solve for λ to get:

$$\hat{\lambda} = k \Big/ \sum u_i$$

Equivalently, the mean time to failure is:

$$1 \Big/ \hat{\lambda} = \sum u_i / k$$

This differs from the standard MLE for the exponential distribution in that the any censored observations are considered only in the numerator.

Statistical inference based on survival data

Parametric approaches are used either when a suitable model or distribution is fitted to the survival data or when a distribution can be assumed for the population from which the sample is known. Commonly used survival distributions are Exponential, Weibull, Log-normal and Gamma. If a survival distribution is formed to fit the data properly. The survival pattern can be described by the parameters. In a compact way and the statistical inference can be based on the chosen distribution. If the search for an appropriate model or distribution is too time consuming or not economical or no theoretical distribution fits the data non-parametric methods should be considered in this case.

Functions of survival Time

The life distribution of survival times is usually described or characterized by 3 functions.

- Survivor Function, denoted by S(t)
- Hazard Function, denoted by h(t)
- Probability Density Function, denoted by f(t)

A basic problem in survival data analysis to estimate from the sampled data one or more of these three functions and to draw inference about a survival pattern in the population.

Survivor Function

Let T denote the survival time. The distribution of T can be characterized by three equivalent functions.

This function denoted by S(t), is defined as the probability that an individual survives longer than t.

S(t) = P (an individual survives longer than t)

= P(T >t)

From the definition of the cumulative distribution function F(t) of T,

S(t) = 1-P (an individual fails before time t)

= 1 – P(T < t) = 1 - F(t)

Here S(t) is a non-increasing function of time t with the properties

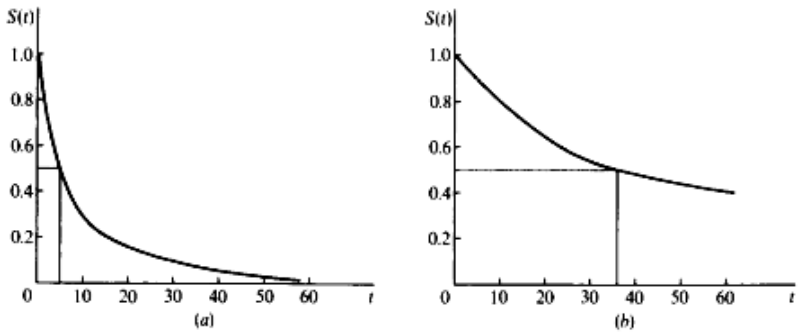
$$S(t) = \begin{cases} 1 & for\, t = 0 \\ 0 & for\, t = \infty \end{cases}$$

That is, the probability of surviving at least at the time zero is 1 and that of surviving an infinite time is zero. The function S(t) is also known as the cumulative survival rate. The graph of S(t) is called the survival curve.

The survivorship function is estimated as the proportion of patients surviving longer than t,

$$\hat{S}(t) = \frac{Number\, of\ \ patients\, surviving\, longer\, than\, t}{Total\, number\, of\ \ patients} \qquad \rightarrow (1)$$

For example, consider the following set of survival data: 4, 6, 6+, 10+, 15and 20.We can compute S(5) = 5/6= 0.8339



Two examples of survival curves.

a, represents low survival rate or short survival time.

b, represents high survival rate or longer survival.

When censored observations are present, in equation (1) is no longer appropriate for estimating S(t). Nonparametric methods of estimating S(t) should be used.

Probability Density Function

The survival time T has a probability density function defined as the limit of the probability that an individual fails in the short interval t to t +Δt per unit width Δt, or simply the probability of failure in a small interval per unit time. It can be expressed as

$$f(t) = \frac{\lim_{\Delta t \to 0} P[an\, individual\, dying\, in\, the \text{int} erval(t, t + \Delta t)]}{\Delta t}$$

The graph of f (t) is called the density curve. Figure a and b give two examples of the density curve. The density function has the following two properties:

1. f (t) is a nonnegative function:

$$f(t) \geq 0 \qquad \text{for all } t \geq 0$$

$$= 0 \qquad \text{for } t<0$$

2. The area between the density curve and the t axis is equal to 1.

If there are no censored observations, the probability density function *f(t)* is estimated as the proportion of patients dying in an interval per unit width:

$$\hat{f}(t) = \frac{No.of\ \ patents dying\ in the Interval begining\ at time t}{(Total\ number of\ \ patients) \times (Interval\ width)} \qquad \rightarrow (2)$$

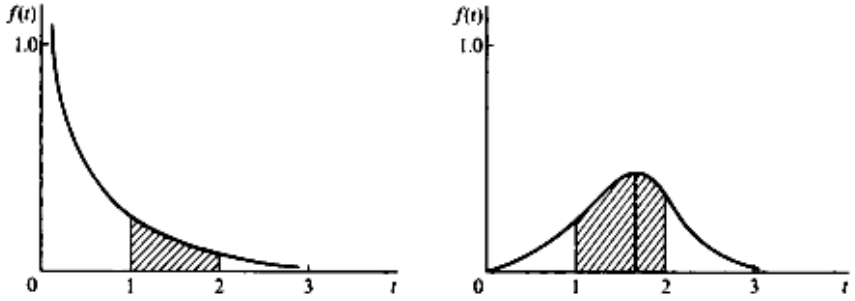


Figure (*a*) and (*b*) **Two examples of density curves.**

Similar to the estimation of S(t), when censored observations are present, in equation (2). We need to use non-parametric methods for estimating *f(t)*.

Hazard Function

The hazard function h(t) of survival time T gives the conditional failure rate. This is defined as the probability of failure during a very small time interval, assuming that the individual has survived to the beginning of the interval, or as the limit of the probability that an individual fails in a very short interval, t + Δt, given that the individual has survived to time t:

$$h(t) = \frac{\lim_{\Delta t \to 0} P\left[\begin{array}{l} an\ individual\ \ fails\ in the time\ Interval\ (t, t+\Delta t) \\ given the individual\ has\ survived\ to t \end{array}\right]}{\Delta t}$$

The hazard function is also known as the instantaneous failure rate, force of mortality, conditional mortality rate, and age-specific failure rate.

When there are no censored observations the hazard function is estimated as the proportion of patients dying in an interval per unit time, given that they have survived to the beginning of the interval:

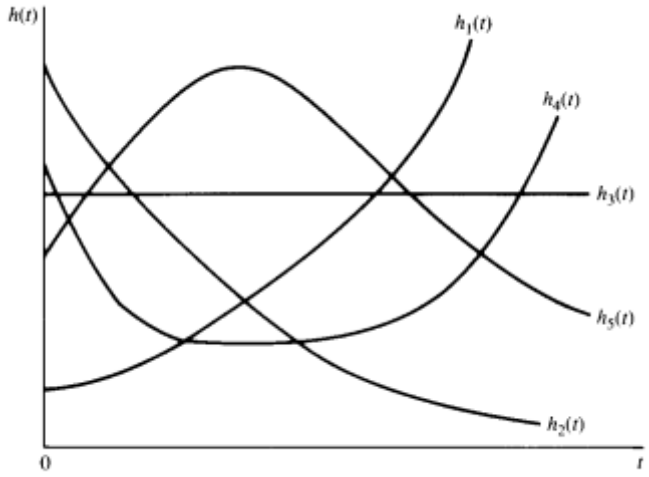
$$\hat{h}(t) = \frac{No.of\ \ patents dying\ in the Interval begining\ at time t}{(Total number of\ \ patients surviving at time t) \times (Interval\ width)}$$

$$= \frac{No.of\ \ patents dying\ \ per unit time in the Interval}{number of\ \ patients surviving at t}$$

Actuaries usually use the average hazard rate of the interval in which the number of patients dying per unit time in the interval is divided by the average number of survivors at the midpoint of the interval:

$$\hat{h}(t) = \frac{No.of\ \ patents dying\ \ per unit time in the Interval}{(number of\ \ patients surviving at time t) - (number of\ \ deaths in the Interval)/2}$$

The hazard function may increase, decrease, remain constant, or indicate a more complicated process.



Examples of the hazard function.

Example: Functions of survival Time

The survival data of 40 patients with myeloma. The survival times are grouped into intervals of five months. Find the number of patients dying in interval. Compute and plot the estimated survivorship function, density function, and hazard function and Plot them.

Survival Time t (months)	***0-5***	***5-10***	***10-15***	***15-20***	***20-25***	***25-30***
no. of patients surviving at beginning of interval	40	35	28	22	18	13
Survival Time t (months)	***30-35***	***35-40***	***40-45***	***45-50***	***≥ 50***	-
no. of patients surviving at beginning of interval	9	5	5	3	2	-

Procedure

- To calculate the number of patients dying in interval.
- To calculate the estimated survivorship function,

$$\hat{S}(t) = \frac{Number\ of\ \ patients\ surviving\ longer\ than\ t}{Total\ number\ of\ \ patients}$$

- To calculate the estimated density function,

$$\hat{f}(t) = \frac{No.of\ \ patents\ dying\ in\ the\ Interval\ begining\ at\ time\ t}{(Total\ number\ of\ \ patients) \times (Interval\ width)}$$

- To calculate the estimated hazard function,

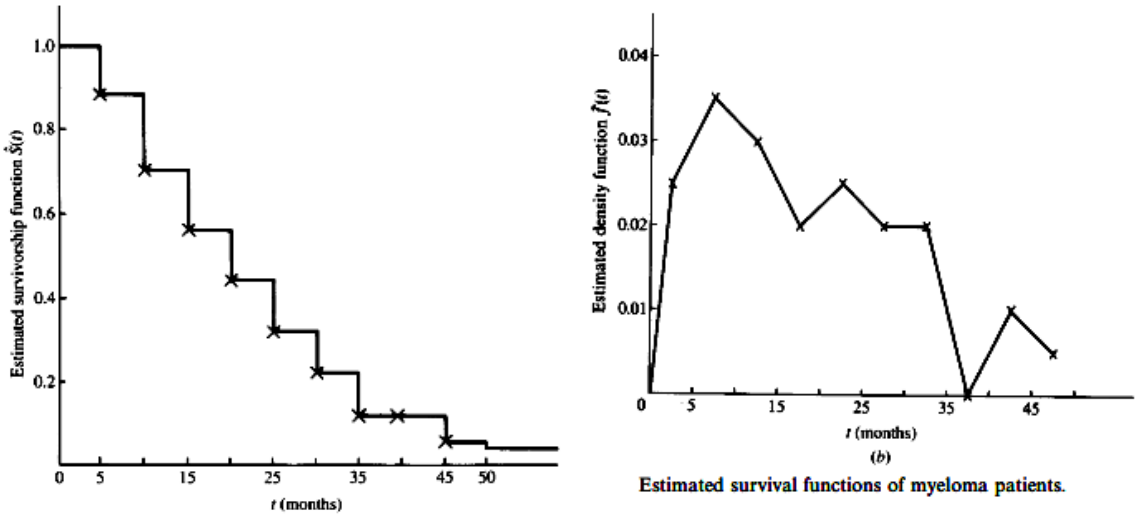
$$\hat{h}(t) = \frac{No.of\ \ patents\ dying\ \ per\ unit\ time\ in\ the\ Interval}{(number\ of\ \ patients\ surviving\ at\ time\ t) - (number\ of\ \ deaths\ in\ the\ Interval)/2}$$

Calculation

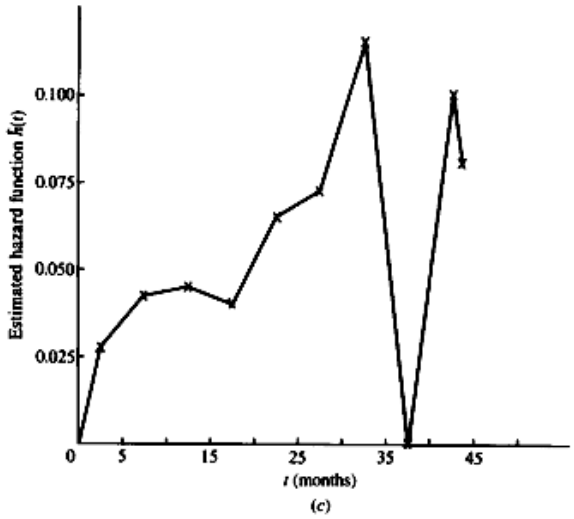
Calculate estimated survival time,

Survival data and estimated survival functions of 40 myeloma patients

Survival Time t (months)	no. of patients surviving at beginning of interval	No. of patients dying in the interval	$\hat{S}(t)$	$\hat{f}(t)$	$\hat{h}(t)$
0-5	40	5	1.000	0.025	0.027
5-10	35	7	0.875	0.035	0.044
10-15	28	6	0.700	0.030	0.048
15-20	22	4	0.550	0.020	0.040
20-25	18	5	0.450	0.025	0.065
25-30	13	4	0.325	0.020	0.072
30-35	9	4	0.225	0.020	0.114
35-40	5	0	0.125	0.000	0.000
40-45	5	2	0.125	0.010	0.100
45-50	3	1	0.075	0.005	0.080
≥50	2	2	0.050	-	-



Estimated survival functions of myeloma patients.



Result

The median survival time of myeloma patients is approximately 17.5 months, and the peak of high frequency of death occurs in 5 to 10 months. In addition, the hazard function shows an increasing trend and reaches its peak at approximately 32.5 months and then fluctuates.

Cumulative hazard function

It is defined as,

$$H(t) = \int_0^t h(x)dx$$

Where H(t) is cumulative hazard function t time t.

Relationships of survival functions

1. Hazard function

$$h(t) = \frac{f(t)}{S(t)} \qquad \rightarrow (1)$$

2. Since the probability density function is the derivative of the cumulative distribution function,

$$f(t) = \frac{d}{dt}[1 - S(t)] = -S'(t) \qquad \rightarrow (2)$$

3. Substituting (2) into (1) yields,

$$h(t) = -\frac{S'(t)}{S(t)} = -\frac{d}{dt}\log S(t) \qquad \rightarrow (3)$$

4. Integrating (3) from zero to t and using S(0) = 1, we have

$$-\int_0^t h(x)dx = \log S(t)$$

$$H(t) = -\log S(t)$$

$$S(t) = \exp[-H(t)] = \exp\left[-\int_0^t h(x)dx\right] \quad \rightarrow (4)$$

From (1) and (4) we obtain

$$f(t) = h(t)\exp[-H(t)] \quad \rightarrow (5)$$

5. Using the definition of the cumulative distribution function,

$$F(t) = \int_0^t f(x)dx$$

$$1\text{- } F(t) = S(t)$$

Survival distribution and their properties

Exponential distribution

The one-parameter exponential distribution has the following density function:

f(t)= $\lambda e^{-\lambda t}$ for λ> 0 (scale parameter)

Denoted T~ Exp(λ). For t > 0,

F(t) = 1-$e^{-\lambda t}$ S(t) = $e^{-\lambda t}$

The hazard function is

h(t) = λ (constant hazard function)

The cumulative hazard function is

H(t) = λt

Note:

For exponential distribution of the form **$f(t) = (1/\theta)\times e^{-(t/\theta)}$,** replace λ by (1/θ) in S(t), h(t) and H(t).

Gamma distribution

A continuous random variable t is said to have a gamma distribution with parameters r>0 and λ>0, shown as X ~ Gamma(r,λ), if its PDF is given by

$$f(t) = \frac{\lambda}{\Gamma(\gamma)}(\lambda t)^{\gamma-1}e^{-\lambda t} \qquad t > 0,\ \gamma > 0,\ \lambda > 0 \qquad \rightarrow (1)$$

For the Erlangian distribution, it can be shown that

$$F(t) = 1 - \sum_{k=0}^{n-1}\frac{e^{-\lambda t}(\lambda t)^k}{k!} \qquad \rightarrow (2)$$

Thus, the survivorship function 1 - $F(t)$ is

$$S(t) = \int_t^{\infty}\frac{\lambda}{\Gamma(\gamma)}(\lambda x)^{\gamma-1}\ e^{-\lambda x}dx \qquad \rightarrow (3)$$

For the gamma distribution

$$S(t) = e^{-1}\sum_{k=0}^{n=1}\frac{(\lambda t)^k}{k!} \qquad \rightarrow (4)$$

for the Erlangian distribution.

Since the hazard function is the ratio of $f(t)$ to $S(t)$, it can be calculated from(1) and (4). When γ is an integer n,

$$h(t) = \frac{\lambda(\lambda t)^{n-1}}{(n-1)!\sum_{k=0}^{n-1}(\frac{1}{k!})(\lambda t)^k}$$

Weibull distribution

The one-parameter Weibull distribution has the following density function:

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

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

The hazard function is

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

The cumulative hazard function is

$$H(t) = (\lambda t)^p$$

Log-normal distribution

The log-normal distribution is denoted LN(μ,σ^2) ~ exp{N(μ,σ^2)}: It is defined as follows:

$$f(t) = \frac{1}{t\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2\sigma^2}(\log t - \mu)^2\right] \qquad t > 0,\ \sigma > 2 \qquad (1)$$

and

$$S(t) = \frac{1}{\sigma\sqrt{2\pi}}\int_t^{\infty}\frac{1}{x}\exp\left[-\frac{1}{2\sigma^2}(\log x - \mu)^2\right]dx \qquad (2)$$

Let a= exp(-μ). Then – $\mu = loga$,

$$f(t) = \frac{1}{t\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2\sigma^2}(\log at)^2\right] \qquad (3)$$

and

$$S(t) = \frac{1}{\sigma\sqrt{2\pi}}\int_t^{\infty}\frac{1}{x}\exp\left[-\frac{1}{2\sigma^2}(\log x - \mu)^2\right]dx \qquad (4)$$

$$=1\text{-}G\left(\log\frac{at}{\sigma}\right)\rightarrow (5)$$

$$h(t) = \frac{\left(\frac{1}{t\sigma\sqrt{2\pi}}\right)\exp[-\log at]^2)}{1-G(\log at/\sigma)}$$

One-Parameter Exponential Distribution

The one-parameter exponential distribution has the following density function;

$$f(t) = \lambda e^{-\lambda t} \qquad (1)$$

survivorship function;

$$S(t) = e^{-\lambda t} \qquad (2)$$

and hazard function;

$$h(t) = \lambda \qquad (3)$$

Where, $t \geq 0$, $\lambda > 0$. Obviously, the exponential distribution is characterized by one parameter, λ. The estimation of λ by maximum likelihood methods for data without censored observations will be given first followed by the case with censored observations.

Estimation of λ for Data without Censored Observations

Suppose that there are *n* persons in the study and everyone is followed to death or failure. Let $t_1, t_2, \ldots, t_n$ be the exact survival times of the *n* people. The likelihood function, using (1) and $l(b) = \log L(b) = \sum_{i=1}^{r} \log\left[f(t_i, b)\right] + \sum_{i=r+1}^{n} \log\left[S(t_i^+, b)\right]$, is

$$L = \prod_{i=1}^{n} \lambda e^{-\lambda t_i}$$

and the log-likelihood function is,

$$l(\lambda) = n \log \lambda - \lambda \sum_{i=1}^{n} t_i \qquad (4)$$

From $\frac{\partial l(b)}{\partial b_j} = 0,\ j = 1, 2, \ldots, p$, the MLE of λ is

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} t_i} \qquad (5)$$

Since the mean μ of the exponential distribution is 1/λ and a MLE is invariant under an one-to-one transformation, the MLE of μ is

$$\hat{\mu} = \frac{1}{\hat{\lambda}} = \frac{\sum_{i=1}^{n} t_i}{n} = \bar{t} \qquad (6)$$

It can be shown $2n\hat{\mu}/\mu$ has an exact chi-square distribution with 2*n* degrees of freedom. Since λ = 1/μ and $\hat{\lambda} = 1/\hat{\mu}$, an exact 100(1 - α)% confidence interval for λ is

$$\frac{\hat{\lambda}\chi^2_{2n,1-\alpha/2}}{2n} < \lambda < \frac{\hat{\lambda}\chi^2_{2n,\alpha/2}}{2n} \qquad (7)$$

Where $\chi^2_{2n,\alpha}$ is the 100α percentage point of the chi-square distribution with 2n degrees of freedom, that is, $P(\chi^2_{2n} > \chi^2_{2n,\alpha}) = \alpha$. When n is large ($n \geq 25$), $\hat{\lambda}$ is approximately normally distributed with mean λ and variance λ^2/n. Thus, an approximate 100(1 - α)% confidence interval for λ is

$$\hat{\lambda} - \frac{\hat{\lambda} Z_{\alpha/2}}{\sqrt{n}} < \lambda < \hat{\lambda} + \frac{\hat{\lambda} Z_{\alpha/2}}{\sqrt{n}} \qquad (8)$$

Where $Z_{\alpha/2}$ is the 100α/2 percentage point, $P(Z > Z_{\alpha/2})$, of the standard normal distribution.

Since $2n\hat{\mu}/\mu$ has an exact chi-square distribution with 2n degrees of freedom, an exact 100(1 - α)% confidence interval for the mean survival time is

$$\frac{2n\hat{\mu}}{\chi^2_{2n,\alpha/2}} < \mu < \frac{2n\hat{\mu}}{\chi^2_{2n,1-\alpha/2}} \qquad (9)$$

Example

In a study of a new insecticide, 20 insects are exposed. Survival times in seconds are 3, 5, 6, 7, 8, 9, 10, 10, 12, 15, 15, 18, 19, 20, 22, 25, 28, 30, 40, and 60. Assuming that they follow the one-parameter exponential distribution, obtain: (a) The MLE of λ, (b) The MLE of μ, (c) The 95% confidence intervals for λ and μ, and (d) Find the probability of remission for atleast 20 weeks.

Procedure

- To calculate the MLE of relapse rate λ is

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} t_i}$$

- To calculate the MLE of remission time μ is

$$\hat{\mu} = \frac{1}{\hat{\lambda}} = \frac{\sum_{i=1}^{n} t_i}{n} = \bar{t}$$

- To calculate the confidence interval for relapse rate λ is

$$\frac{\hat{\lambda}\chi^2_{2n,1-\alpha/2}}{2n} < \lambda < \frac{\hat{\lambda}\chi^2_{2n,\alpha/2}}{2n}$$

- To calculate the confidence interval for remission time μ is

$$\frac{2n\hat{\mu}}{\chi^2_{2n,\alpha/2}} < \mu < \frac{2n\hat{\mu}}{\chi^2_{2n,1-\alpha/2}}$$

- To calculate the probability of remission time μ is

$$S(t) = e\exp[-\lambda t]$$

Calculation

Here, n = 21, $\sum t_i = 198$

(a) The MLE of λ

The MLE of $\hat{\lambda}$ is

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} t_i}$$

$$\hat{\lambda} = \frac{21}{198} = 0.106 \; per \; week$$

(b) The MLE of μ

The MLE of $\hat{\mu}$ is

$$\hat{\mu} = \frac{1}{\hat{\lambda}} = \frac{\sum_{i=1}^{n} t_i}{n} = \bar{t}$$

$$\hat{\mu} = \frac{198}{21} = 9.429 \, weeks.$$

(c) The 95% confidence intervals for λ and μ

A 95% confidence interval for λ is approximately

$$\frac{\hat{\lambda}\chi^2_{2n,1-\alpha/2}}{2n} < \lambda < \frac{\hat{\lambda}\chi^2_{2n,\alpha/2}}{2n}$$

$$\frac{(0.106)(24.433)}{42} < \lambda < \frac{(0.106)(59.342)}{42}$$

$$0.062 < \lambda < 0.150$$

A 95% confidence interval for λ is 0.062 to 0.150.

A 95% confidence interval for μ is

$$\frac{2n\hat{\mu}}{\chi^2_{2n,\alpha/2}} < \mu < \frac{2n\hat{\mu}}{\chi^2_{2n,1-\alpha/2}}$$

$$\frac{(42)(9.429)}{59.342} < \mu < \frac{(42)(9.429)}{24.433}$$

$$6.673 < \mu < 16.208$$

A 95% confidence interval for μ is 6.673 to 16.208.

(d) The probability of remission for atleast 20 weeks

The probability of staying in remission for at least 20 weeks is,

$$S(t) = e^{-\lambda t}$$

$$\hat{S}(20) = \exp[-0.106(20)] = 0.120.$$

Result

- The MLE of the relapse rate $\hat{\lambda} = 0.106$
- The mean remission time $\hat{\mu} = 9.429$
- A 95% confidence interval for the relapse rate λ is 0.062 to 0.150.
- A 95% confidence interval for the mean remission time μ is 6.673 to 16.208.
- The probability of staying in remission for at least 20 weeks is $\hat{S}(20) = 0.120$.

Estimation of λ for Data with Censored Observations

Suppose that without loss of generality, the study or experiment begins at time 0 with a total of *n* subjects. Survival times are recorded and the data become available when the subjects die one after the other in such a way that the shortest survival time comes first, the second shortest second, and so on. Suppose that the investigator has decided to terminate the study after *r* out of the *n* subjects has died and to sacrifice the remaining *n-r* subjects at that time. Then the survival times for the *n* subjects are

$$t_{(1)} \le t_{(2)} \le \ldots \le t_{(r)} = t^{+}_{(r+1)} = \ \ldots = \ t^{+}_{(n)}$$

where a superscript plus indicates a sacrificed subject, and thus $t^{+}_{(i)}$ is a censored observation. In this case, *n* and *r* are fixed values and all of the *n-r* censored observations are equal.

The likelihood function, using (1), (2) and $l(b) = \log L(b) = \sum_{i=1}^{r} \log\left[f(t_i, b)\right] + \sum_{i=r+1}^{n} \log\left[S(t_i^{+}, b)\right]$ is,

$$L = \frac{n!}{(n-r)!} \prod_{i=1}^{r} \lambda e^{-\lambda t_{(i)}} \left(e^{-\lambda t_{(r)}}\right)^{n-r}$$

And from $\frac{\partial \mathrm{l(b)}}{\partial \mathrm{b_j}} = 0,\ \mathrm{j} = 1, 2, \ldots\ldots, \mathrm{p}$, the MLE of λ is

$$\hat{\lambda} = \frac{r}{\sum_{i=1}^{r} t_{(i)} + \sum_{i=r+1}^{n} t^{+}_{(i)}} \qquad (10)$$

The mean survival time μ= 1/λ can then be estimated by

$$\hat{\mu} = \frac{1}{\hat{\lambda}} = \frac{\sum_{i=1}^{r} t_{(i)} + \sum_{i=r+1}^{n} t^{+}_{(i)}}{r} \qquad (11)$$

It is shown that $2r\lambda/\hat{\lambda}$ has a chi-square distribution with 2*r* degrees of freedom. The mean and variance of $\hat{\lambda}$ are $r\lambda/(r\text{-}1)$ and $\lambda^2/(r\text{-}1)$, respectively.

The 100 (1-α)% confidence interval for λ is

$$\frac{\hat{\lambda}\chi^2_{2r,1-\alpha/2}}{2r} < \lambda < \frac{\hat{\lambda}\chi^2_{2r,\alpha/2}}{2r} \qquad (12)$$

When n is large, the distribution of $\hat{\lambda}$ is approximately normal with mean λ and variance $\lambda^2/(r\text{-}1)$. An approximate 100/(1-α)% confidence interval for λ is

$$\hat{\lambda} - \frac{\hat{\lambda} Z_{\alpha/2}}{\sqrt{r-1}} < \lambda < \hat{\lambda} + \frac{\hat{\lambda} Z_{\alpha/2}}{\sqrt{r-1}} \qquad (13)$$

Epstein and Sobel (1953) show that $2r\hat{\mu}/\mu$ has a chi-square distribution with 2r degrees of freedom. Thus a 100/(1-α)% confidence interval for μ is

$$\frac{2r\hat{\mu}}{\chi^2_{2r,\alpha/2}} < \mu < \frac{2r\hat{\mu}}{\chi^2_{2r,1-\alpha/2}} \qquad (14)$$

They also develop test procedures for the hypothesis $H_0:\mu = \mu_0$ against the alternative $H_1:\mu < \mu_0$. One of their rules of action is to accept H_0 if $\hat{\mu} > c$ and reject H_0 if $\hat{\mu} < c$, where $c = (\mu_0\chi^2_{2r,\alpha})/2r$ and α is the significance level. Or if the estimated mean survival time calculated from (11) is greater then c, the hypothesis H_0 is rejected at the α level.

Example

Suppose that in a laboratory experiment 10 mice are exposed to carcinogens. The experimenter decides to terminate the study after half of the mice are dead and to sacrifice the other half at that time. The survival times of the five dead mice are 4, 5, 8, 9, and 10 weeks. The survival data of the 10mice are 4, 5, 8, 9, 10, 10+, 10+, 10+, 10+, and 10+. Assuming that the failure of these mice follows an exponential distribution. Obtain: (a) The MLE of λ, (b) The MLE of μ, (c) The 95% confidence intervals for λ and μ, and (d) Find the probability that the mice will survive longer than 8 weeks.

Procedure

- To calculate the MLE of relapse rate λ is

$$\hat{\lambda} = \frac{r}{\sum_{i=1}^{r} t_{(i)} + \sum_{i=r+1}^{n} t_{(i)}^{+}}$$

- To calculate the MLE of remission time μ is

$$\hat{\mu} = \frac{1}{\hat{\lambda}} = \frac{\sum_{i=1}^{r} t_{(i)} + \sum_{i=r+1}^{n} t_{(i)}^{+}}{r}$$

- To calculate the confidence interval for relapse rate λ is

$$\frac{\hat{\lambda}\chi^2_{2r,1-\alpha/2}}{2r} < \lambda < \frac{\hat{\lambda}\chi^2_{2r,\alpha/2}}{2r}$$

- To calculate the confidence interval for remission time μ is

$$\frac{2r\hat{\mu}}{\chi^2_{2r,\alpha/2}} < \mu < \frac{2r\hat{\mu}}{\chi^2_{2r,1-\alpha/2}}$$

- To calculate the probability of remission time μ is

$$S(t) = e\exp[-\lambda t]$$

Calculation

Here, r = 5, $\sum_{i=1}^{r} t_{(i)} = 36$ and $\sum_{i=r+1}^{n} t_{(i)}^{+} = 50$

(a) The MLE of λ

The MLE of λ is,

$$\hat{\lambda} = \frac{r}{\sum_{i=1}^{r} t_{(i)} + \sum_{i=r+1}^{n} t_{(i)}^{+}}$$

$$\hat{\lambda} = \frac{5}{36+50} = 0.058 \, per \, week$$

(b) The MLE of μ

The MLE of μ is

$$\hat{\mu} = \frac{1}{\hat{\lambda}} = \frac{\sum_{i=1}^{r} t_{(i)} + \sum_{i=r+1}^{n} t_{(i)}^{+}}{r}$$

$$\hat{\mu} = \frac{1}{0.058} = 17.241 weeks$$

(c) The 95% confidence intervals for λ and μ

A 95% confidence interval for λ is approximately

$$\frac{\hat{\lambda}\chi^2_{2r,1-\alpha/2}}{2r} < \lambda < \frac{\hat{\lambda}\chi^2_{2r,\alpha/2}}{2r}$$

$$\frac{(0.058)(3.247)}{10} < \lambda < \frac{(0.058)(20.483)}{10}$$

$$0.019 < \lambda < 0.119$$

A 95% confidence interval for λ is 0.019 to 0.119.

A 95% confidence interval for μ is

$$\frac{2r\hat{\mu}}{\chi^2_{2r,\alpha/2}} < \mu < \frac{2r\hat{\mu}}{\chi^2_{2r,1-\alpha/2}}$$

$$\frac{(2)(5)(17.241)}{20.483} < \mu < \frac{(2)(5)(9.429)}{3.247}$$

$$8.417 < \mu < 53.098$$

A 95% confidence interval for μ is 8.417 to 53.098.

(d) The probability that the mice will survive longer than 8 weeks

The probability that the mouse exposed to the same carcinogen will survive longer than 8 weeks is,

$$S(t) = e^{-\lambda t}$$

$$\hat{S}(8) = \exp[-0.058(8)] = 0.629.$$

The probability of dying in 8weeks is then 1-0.629 = 0.371.

Result

- The MLE of $\hat{\lambda}$ = 3.247
- The MLE of $\hat{\mu}$ = 17.241
- A 95% confidence interval for λ is 0.019 to 0.119.
- A 95% confidence interval for μ is 8.417 to 53.098.
- The probability that the mice will survive longer than 8 weeks is $\hat{S}(8) = 0.629$
- The probability of dying in 8weeks is 0.371.

Estimation of λ for Data with slightly different situation may arise

A slightly different situation may arise in laboratory experiments. Instead of terminating the study after the r^{th} death, the experimenter may stop after aperiod of time T, which may be six months or a year. If we denote the number of deaths between 0 and T as r, the survival data may look as follows:

$$t_{(1)} \leq t_{(2)} \leq \ldots \leq t_{(r)} = t^{+}_{(r+1)} = \ \ldots = \ t^{+}_{(n)} = T$$

Mathematical derivations of the MLE of λ and μ are exactly the same and equation (10) can be used.

Suppose that the study begins at time 0 and terminates at time *T* and there are a total of *n* people entered. Let *r* be the number of patients who die before or at time *T* and *n-r* the number of patients who are lost to follow-up during the study period or remain alive at time *T*. The data look as follows: $t_1, t_2, \ldots, t_r, t^{+}_{r+1}, \ldots, t^{+}_{n}$. Ordering the r uncensored observations according to their magnitude, we have

$$t_{(1)} \leq t_{(2)} \leq \ldots \leq t_r, \ t^{+}_{r+1}, \ldots, t^{+}_{n}$$

The likelihood function, using (1), (2) and $l(b) = \log L(b) = \sum_{i=1}^{r} \log\left[f(t_i, b)\right] + \sum_{i=r+1}^{n} \log\left[S(t_i^{+}, b)\right]$, is

$$L = \prod_{i=1}^{r} \lambda e^{-\lambda t_{(i)}} \prod_{i=r+1}^{n} \lambda e^{-\lambda t_i^{+}}$$

and the log-likelihood function is,

$$l(\lambda) = n\lambda - \lambda \sum_{i=1}^{r} t_i - \lambda \sum_{i=r+1}^{n} t_i^{+} \qquad (15)$$

From $\frac{\partial l(b)}{\partial b_j} = 0, \ j = 1, 2, \ldots., p$, the MLE of the parameter λ is

$$\hat{\lambda} = \frac{r}{\sum_{i=1}^{r} t_{(i)} + \sum_{i=r+1}^{n} t_i^+} \tag{16}$$

The mean survival time μ= 1/λ can then be estimated by,

$$\hat{\mu} = \frac{1}{\hat{\lambda}} = \frac{\sum_{i=1}^{r} t_{(i)} + \sum_{i=r+1}^{n} t_i^+}{r} \tag{17}$$

The sum of all of the observations, censored and uncensored, divided by the number of uncensored observations, gives the MLE of the mean survival time. To overcome the mathematical difficulties arising when all of the observations are censored *(r=0)*,

$$\hat{\mu} = \sum_{i=1}^{n} t_i^+ \tag{18}$$

The distribution of $\hat{\lambda}$ for large *n* is approximately normal with mean λ and variance:

$$Var(\hat{\lambda}) = \frac{\lambda^2}{\sum_{i=1}^{n} \left(1 - e^{-\lambda T_i}\right)} \tag{19}$$

Where T_i is the time that the i^{th} person is under observation. In other words, T_i is computed from the time the i^{th} person enters the study to the end of the study. If the observation times T_i are not known, the following quick estimate of $Var(\hat{\lambda})$ can be used:

$$\widehat{Var}(\hat{\lambda}) = \frac{\hat{\lambda}^2}{r} \tag{20}$$

Thus an approximate 100(1 - λ)% confidence interval for λ is, by $\hat{b}_i - Z_{\alpha/2}\sqrt{v_{ii}},\ \hat{b}_i + Z_{\alpha/2}\sqrt{v_{ii}}$

$$\hat{\lambda} - Z_{\alpha/2}\sqrt{\widehat{Var}(\hat{\lambda})} < \lambda < \hat{\lambda} + Z_{\alpha/2}\sqrt{\widehat{Var}(\hat{\lambda})} \tag{21}$$

The distribution of $\hat{\mu}$ is approximately normal with mean λ and variance:

$$Var(\hat{\mu}) = \frac{\mu^2}{\sum_{i=1}^{n} \left(1 - e^{-\lambda T_i}\right)} \tag{22}$$

Again quick estimate of $Var(\hat{\mu})$ is

$$\widehat{Var}(\hat{\mu}) = \frac{\hat{\mu}^2}{r} \tag{23}$$

Thus an approximate 100(1 - α)% confidence interval for μ is, by $\hat{b}_i - Z_{\alpha/2}\sqrt{v_{ii}},\ \hat{b}_i + Z_{\alpha/2}\sqrt{v_{ii}}$

$$\hat{\mu} - Z_{\alpha/2}\sqrt{\widehat{Var}(\hat{\mu})} < \mu < \hat{\mu} + Z_{\alpha/2}\sqrt{\widehat{Var}(\hat{\mu})} \tag{24}$$

Example

Consider the remission duration of the 21 leukemia patients receiving 6-MP the remission times in weeks were 6, 6, 6, 7, 10, 13, 16, 22, 23, 6+, 9+, 10+, 11+, 17+, 19+, 20+, 25+, 32+, 32+, 34+ and 35+. Assume that the remission time follows on exponential distribution. To estimate the MLEs of the relapse rate and the mean remission time and its confidence interval and find the probability of staying in remission for one year (52 weeks) or more.

Procedure

- To calculate the MLE of relapse rate λ is

$$\hat{\lambda} = \frac{r}{\sum_{i=1}^{r} t_{(i)} + \sum_{i=r+1}^{n} t_{(i)}^{+}}$$

- To calculate the MLE of remission time μ is

$$\hat{\mu} = \frac{1}{\hat{\lambda}} = \frac{\sum_{i=1}^{r} t_{(i)} + \sum_{i=r+1}^{n} t_{(i)}^{+}}{r}$$

- To calculate the Variance of λ is

$$\widehat{Var}(\hat{\lambda}) = \frac{\hat{\lambda}^2}{r}$$

- To calculate the Variance of μ is

$$\widehat{Var}(\hat{\mu}) = \frac{\hat{\mu}^2}{r}$$

- To calculate the confidence interval for relapse rate λ is

$$\hat{\lambda} - Z_{\alpha/2}\sqrt{\widehat{Var}(\hat{\lambda})} < \lambda < \hat{\lambda} + Z_{\alpha/2}\sqrt{\widehat{Var}(\hat{\lambda})}$$

- To calculate the confidence interval for remission time μ is

$$\hat{\mu} - Z_{\alpha/2}\sqrt{\widehat{Var}(\hat{\mu})} < \mu < \hat{\mu} + Z_{\alpha/2}\sqrt{\widehat{Var}(\hat{\mu})}$$

- To calculate the probability of remission time μ is

$$S(t) = e\exp[-\lambda t]$$

Calculation

Here, r = 9, $\sum_{i=1}^{r} t_{(i)} = 109$ and $\sum_{i=r+1}^{n} t_{(i)}^{+} = 250$

The MLE of λ is,

$$\hat{\lambda} = \frac{9}{109 + 250} = 0.025 \, per \, week$$

The MLE of μ is

$$\hat{\mu} = \frac{1}{0.025} = 40 \, weeks$$

Variance of λ is

$$\widehat{Var}(\hat{\lambda}) = \frac{(0.025)^2}{9} = 0.000069$$

Variance of μ is

$$\widehat{Var}(\hat{\mu}) = \frac{(40)^2}{9} = 177.78$$

A 95% confidence interval for λ is approximately

$$0.025 - 1.96\sqrt{0.000069} < \lambda < 0.025 + 1.96\sqrt{0.000069}$$

$$0.009 < \lambda < 0.041$$

A 95% confidence interval for λ is 0.009 to 0.041.

A 95% confidence interval for μ is

$$40 - 1.96\sqrt{177.78} < \mu < 40 + 1.96\sqrt{177.78}$$

$$13.867 < \mu < 66.133$$

A 95% confidence interval for μ is 13.867 to 66.133.

The probability of staying in remission for one year (52 weeks) or more is,

$\hat{S}(52)$ = exp[-0.025(52)] = 0.273.

Result

- The MLE of $\hat{\lambda}$ = 0.025
- The MLE of $\hat{\mu}$ = 40
- Variance of λ is 0.000069
- Variance of μ is 177.78

- A 95% confidence interval for λ is 0.009 to 0.041.
- A 95% confidence interval for μ is 13.867 to 66.133.
- The probability of staying in remission for one year (52 weeks) or more is 0.273.

Test based on LR

Let $l_{LL}(\alpha, \gamma)$ and $l_{LL}(\hat{\alpha}, \hat{\gamma})$ denote its log-likelihood function and the log-likelihood with $(\hat{\alpha}, \hat{\gamma})$, the MLE of (α, γ).

1. Testing the hypothesis that the underlying distribution is exponential with known parameter λ_0. The null hypothesis is

   H_0: the underlying distribution is the exponential distribution with $\lambda = \lambda_0$.

   The likelihood ratio test statistic is

   $$X_L = 2[l_E(\hat{\lambda}) - l_E(\lambda_0)] \qquad \rightarrow (1)$$

   Under H_0, X_L has an asymptotic chi-square distribution with 1 degree of freedom. H_0 is rejected if $X_L > X^2_{1,\alpha}$, where α is the significance level.

2. Testing the hypothesis that the underlying distribution is Weibull with known parameter λ_0 and γ_0. The null hypothesis is

   H_0: the underlying distribution is the Weibull distribution with $\lambda = \lambda_0$ and $\gamma = \gamma_0$.

   The likelihood ratio test statistic is

   $$X_L = 2[l_w(\hat{\lambda}, \hat{\gamma}) - l_w(\lambda_0, \gamma_0)] \qquad \rightarrow (2)$$

   Under H_0, X_L has an asymptotic chi-square distribution with 2 degree of freedom. H_0 is rejected if $X_L > X^2_{2,\alpha}$, where α is the significance level.

3. Testing the hypothesis that the underlying distribution is lognormal with known parameter μ_0 and σ_0^2. The likelihood ratio test statistic is

   $$X_L = 2[l_{LN}(\hat{\mu}, \hat{\sigma}^2) - l_{LN}(\mu_0, \sigma_0^2)] \qquad \rightarrow (3)$$

   Under H_0, X_L has an asymptotic chi-square distribution with 2 degree of freedom. H_0 is rejected if $X_L > X^2_{2,\alpha}$, where α is the significance level.

4. Testing the hypothesis that the underlying distribution is Standard Gamma with known parameter λ_0 and γ_0. The likelihood ratio test statistic is

   $$X_L = 2[l_G(\hat{\lambda}, \hat{\gamma}) - l_G(\lambda_0, \gamma_0)] \qquad \rightarrow (4)$$

   Under H_0, X_L has an asymptotic chi-square distribution with 2 degree of freedom.

Example

Consider the following survival times in weeks of 10 mice with a given tumor: 1, 3, 5, 8, 10+, 15, 18, 19, 22 and 25+. Do the data follow the exponential distribution with λ=0.06? Use the log likelihood ratio test.

Procedure

- To calculate the MLE of relapse rate λ is

$$\hat{\lambda} = \frac{r}{\sum_{i=1}^{r} t_{(i)} + \sum_{i=r+1}^{n} t_{(i)}^{+}}$$

- To calculate the log likelihood ratio test statistic

$$l_E(\hat{\lambda}) = \sum_{i=1}^{r} \log(\hat{\lambda} e^{-\lambda t_i}) + \sum_{i=r+1}^{n} \log(e^{-\lambda t_i^{+}}) = r \log \hat{\lambda} - \hat{\lambda} \sum_{i=1}^{r} t_i - \hat{\lambda} \sum_{i=r+1}^{n} t_i^{+}$$

$$X_L = 2[l_E(\hat{\lambda}) - l_E(\lambda_0)]$$

Calculation

Hypothesis

H_0: the underlying distribution of the observed data is exponential with $\lambda_0 = 0.06$.

H_1: the underlying distribution of the observed data is not exponential with $\lambda_0 = 0.06$

Here, $n = 10$, $r = 8$, $\sum_{i=1}^{r} t_i = 91$ and $\sum_{i=1+1}^{n} t_i^{+} = 35$.

The MLE of λ is based on $\hat{\lambda} = \frac{r}{\sum_{i=1}^{r} t_{(i)} + \sum_{i=r+1}^{n} t_i^{+}}$ and

$$l_E(\hat{\lambda}) = \sum_{i=1}^{r} \log(\hat{\lambda} e^{-\lambda t_i}) + \sum_{i=r+1}^{n} \log(e^{-\lambda t_i^{+}}) = r \log \hat{\lambda} - \hat{\lambda} \sum_{i=1}^{r} t_i - \hat{\lambda} \sum_{i=r+1}^{n} t_i^{+}$$

$$\hat{\lambda} = \frac{8}{91 + 35} = 0.0635$$

$$l_E(\hat{\lambda}) = 8(\log 0.0635) - (0.0635 \times 91) - (0.0635 \times 35) = -17.58$$

Under H_0,

$$l_E(\hat{\lambda}_0) = 8(\log 0.06) - (0.06 \times 91) - (0.06 \times 35) = -17.33$$

Thus, the log likelihood ratio test statistic

$$X_L = 2[l_E(\hat{\lambda}) - l_E(\lambda_0)]$$

$$X_L = 2[-17.58 - (-17.33)] = -0.5$$

Table value

$$\chi^2_{1,0.05} = 3.84$$

Therefore, we cannot reject the null hypothesis that the data are from the exponential distribution with $\lambda = 0.06$.

Result

Since the calculated value is less than the table value (i.e., $0.5 < 3.84$), we cannot reject the null hypothesis. Hence we conclude that the data is from the exponential distribution with $\lambda = 0.06$.