SURVIVAL

Although life table analysis may be useful in many differing situations and disciplines, for simplicity, the usual survival-time-to-death terminology will be used here.

Notation

The following notation is used throughout this section unless otherwise stated:

X_j	Time from starting event to terminal event or censoring for case *j*
w_j	Weight for case *j*
k	Total number of intervals
t_i	Beginning time for *i*th interval
h_i	Width of interval *i*
c_i	Sum of weights of cases censored in interval *i*
d_i	Sum of weights of cases experiencing the terminal event in interval *i*

Construction of Life Table (Gehan, 1975)

Computation of Intervals

The widths of the intervals for the actuarial calculations must be defined by the user. In addition to the last interval specified, an additional interval is automatically created to take care of any times exceeding the last. If the upper limits are not in ascending order, a message is printed and the procedure stops. If the interval width does not divide the time range into an integral number of intervals, a warning is printed and the interval width is reset so that the number of intervals will be the nearest integer to that resulting from the user specification.

Count of Events and Censoring

For each case, the interval *i* into which the survival time falls is determined.

$$t_i \le X_j < t_{i+1}$$

If X_j exceeds t_k, the starting time for the last interval, it is included in the last interval. The status code is examined to determine whether the observed time is time to event or time to censoring. If it is time to censoring, that is, the terminal event did not occur, c_i is incremented by the case weight. If it is time to event, d_i is incremented by the case weight.

Calculation of Survival Functions

For each interval, the following are calculated:

Number Alive at the Beginning

$$l_i = l_{i-1} - c_{i-1} - d_{i-1}$$

where l_1 is the sum of weights of all cases in the table.

Number Exposed to Risk of an Event

$$r_i = l_i - c_i/2$$

Proportion Terminating

$$q_i = \frac{d_i}{r_i}$$

Proportion Surviving

$$p_i = 1 - q_i$$

Cumulative Proportion Surviving at End of Interval

$$P_i = P_{i-1}\, p_i$$

where

$$P_0 = 1$$

Probability Density Function

$$f_i = \frac{P_{i-1} - P_i}{h_i}$$

Hazard Rate

$$\lambda_i = \frac{2q_i}{h_i\left(1 + p_i\right)}$$

Standard Error of Probability Surviving

$$\mathrm{se}\left(P_i\right) = P_i \sqrt{\sum_{j=1}^{i} q_j \Big/ \left(r_j\, p_j\right)}$$

Standard Error of Probability Density

$$\text{se}\,(f_i) = \frac{P_i q_i}{h_i}\sqrt{\sum_{j=1}^{i-1} q_j \big/ \left(r_j p_j\right) + p_i \big/ \left(r_i q_i\right)}$$

For the first interval

$$\text{se}\,(f_1) = \frac{P_1 q_1}{h_1}\sqrt{\frac{p_1}{r_1 q_1}}$$

Standard Error of the Hazard Rate

$$\text{se}\,(\lambda_i) = \sqrt{\frac{\lambda_i^2}{r_i q_i}\left\{1 - \left(\frac{\lambda_i h_i}{2}\right)^2\right\}}$$

If $q_i = 0$, the standard error for interval *i* is set to 0.

Median Survival Time

If $P_k > 0.5$ the value printed for median survival time is

$$t_k +$$

Otherwise, let *i* be the interval for which $P_i < 0.5$ and $P_{i-1} \geq 0.5$. The estimate of the median survival time is then

$$Md = \left(t_i\right) + \frac{h_{i-1}\left(P_{i-1} - 0.5\right)}{P_{i-1} - P_i}$$

Comparison of Survival Distributions

The survival times from the groups to be compared are jointly sorted into ascending order. If survival times are equal, the uncensored is taken to be less than the censored. When approximate comparisons are done, they are based on the lifetables, with the beginning of the interval determining the length of survival for cases censored or experiencing the event in that interval.

Notation

The following notation is used throughout this section unless otherwise stated:

N	Number of cases
$X_{(k)}$	Survival time for case *k*, where times are sorted into ascending order so that case 1 has the shortest time and case *N* the longest
w_k	Weight for case *k*
g	Number of nonempty groups in the comparison
W_j	Sum of weights of cases in group *j*
W_c	Sum of weights of censored cases
W_u	Sum of weights of uncensored cases
W	Sum of weights of all cases

Computations

For each case the following are computed:

- ULE_k

  Sum of weights of uncensored cases with survival times less than or equal to that of case *k*.

- CLE_k

  Same as above, but for censored cases.

- UE_k

  Sum of weights of uncensored cases with survival times equal to that of case *k*.

- CE_k

  Same as above, but for censored cases.

The score for case *k* is:

$$S_k = \begin{cases} ULE_k & \text{if } X_k \text{ is censored} \\ A_1 - A_2 - A_3 & \text{if } X_k \text{ is uncensored} \end{cases}$$

where

$A_1 = ULE_k - UE_k$	uncensored cases surviving shorter than case *k*
$A_3 = W_c - CLE_k - CE_k$	censored cases surviving longer than or equal to case *k*
$A_3 = W_u - ULE_k$	uncensored cases surviving longer than case *k*

Test Statistic and Significance (Wilcoxon (Gehan))

The test statistic is

$$D = \frac{(W-1)B}{T}$$

where

$$B = \sum_{j=1}^{g} SS_j^2 \Big/ W_j$$

$SS_j =$ the sum of scores of cases in group *j*

$$T = \sum_{i=1}^{N} S_i^2$$

Under the hypothesis that the groups are samples from the same survival distribution, *D* is asymptotically distributed as a chi square with $(g-1)$ degrees of freedom.

References

Gehan, E. A. 1975. Statistical methods for survival time studies. In: *Cancer Therapy*: *Prognostic Factors and Criteria*. M. J. Staquet, ed. New York: Raven Press, 7–35.

Lee, E., and Desu, M. 1972. A computer program for comparing *k* samples with right censored data. *Computer Programs in Biomedicine*, 2: 315–321.