Title

> **ltable —** Life tables for survival data

Description

ltable displays and graphs life tables for individual-level or aggregate data and optionally presents the likelihood-ratio and log-rank tests for equivalence of groups. ltable also allows you to examine the empirical hazard function through aggregation.

Quick start

Life table for time variable tvar and death indicator died

    ltable tvar died

As above, but graph results with confidence intervals instead and suppress table

    ltable tvar died, graph ci notable

Life tables for each group defined by catvar with results saved to mydata.dta

    ltable tvar died, by(catvar) saving(mydata)

Aggregate time into thirty-day intervals and suppress actuarial adjustment

    ltable tvar died, intervals(30) noadjust

Cumulative failure table for observations where catvar equals 1

    ltable tvar died if catvar==1, failure

Hazard table with frequency weights wvar

    ltable tvar died [fweight=wvar], hazard

Menu

Statistics > Survival analysis > Summary statistics, tests, and tables > Life tables for survival data

Syntax

> ltable *timevar* $[$ *deadvar* $]$ $[$ *if* $]$ $[$ *in* $]$ $[$ *weight* $]$ $[$, *options* $]$

timevar specifies the time of failure or censoring. If *deadvar* is not specified, all values of *timevar* are interpreted as failure times. Observations with *timevar* equal to missing are ignored.

deadvar specifies how the time recorded in *timevar* is to be interpreted. Observations with *deadvar* equal to 0 are treated as censored and all other nonmissing values indicate that *timevar* should be interpreted as a failure time. Observations with *deadvar* equal to missing are ignored.

deadvar does not specify the number of failures. Specify frequency weights for aggregated data recording the number of failures.

options	Description
Main	
notable	display graph only; suppress display of table
graph	present the table graphically, as well as in tabular form
by(*groupvar*)	produce separate tables (or graphs) for each value of *groupvar*
test	report χ^2 measure of differences between groups (2 tests)
overlay	overlay plots on the same graph
survival	display survival table; the default
failure	display cumulative failure table
hazard	display hazard table
ci	graph confidence interval
level(*#*)	set confidence level; default is level(95)
noadjust	suppress actuarial adjustment to the number at risk
tvid(*varname*)	subject ID variable to use with time-varying parameters
intervals(w $\mid$ *numlist*)	time intervals in which data are to be aggregated for tables
saving(*filename* $[$, replace $]$)	save the life-table data to *filename*; use replace to overwrite existing *filename*
Plot	
plotopts(*plot_options*)	affect rendition of the plotted line and plotted points
plot#opts(*plot_options*)	affect rendition of the #th plotted line and plotted points; available only with overlay
CI plot	
ciopts(*rspike_options*)	affect rendition of the confidence intervals
ci#opts(*rspike_options*)	affect rendition of the #th confidence interval; available only with overlay
Add plots	
addplot(*plot*)	add other plots to the generated graph
Y axis, X axis, Titles, Legend, Overall	
twoway_options	any options other than by() documented in [G-3] *twoway_options*
byopts(*byopts*)	how subgraphs are combined, labeled, etc.

plot_options	Description
connect_options	change look of lines or connecting method
marker_options	change look of markers (color, size, etc.)

fweights are allowed; see [U] **11.1.6 weight**.

Options

> Main

notable suppresses displaying the table. This option is often used with graph.

graph requests that the table be presented graphically, as well as in tabular form; when notable is also specified, only the graph is presented. When you specify graph, only one table can be calculated and graphed at a time; see survival, failure, and hazard below.

> graph may not be specified with hazard. Use sts graph to graph estimates of the hazard function.

by(*groupvar*) creates separate tables (or graphs within the same image) for each value of *groupvar*. *groupvar* may be string or numeric.

test presents two χ^2 measures of the differences between groups, the likelihood-ratio test of homogeneity and the log-rank test for equality of survivor functions. The two groups are identified by the by() option, so by() must also be specified.

overlay causes the plot from each group identified in the by() option to be overlaid on the same graph. The default is to generate a separate graph (within the same image) for each group. This option requires the by() option.

survival, failure, and hazard indicate the table to be displayed. If none is specified, the default is the survival table. Specifying failure displays the cumulative failure table. Specifying survival failure would display both the survival and the cumulative failure table. If graph is specified, multiple tables may not be requested.

ci graphs the confidence intervals around survival, failure, or hazard.

level(#) specifies the confidence level, as a percentage, for confidence intervals. The default is level(95) or as set by set level; see [R] **level**.

noadjust suppresses the actuarial adjustment to the number at risk. The default is to consider the adjusted number at risk for each interval as the total at the start minus (the number of censored)/2. If noadjust is specified, the number at risk is simply the total at the start, corresponding to the standard Kaplan–Meier assumption. noadjust should be specified when using ltable to list results corresponding to those produced by sts list; see [ST] **sts list**.

tvid(*varname*) is for use with longitudinal data with time-varying parameters. Each subject appears in the data more than once, and equal values of *varname* identify observations referring to the same subject. When tvid() is specified, only the last observation on each subject is used in making the table. The order of the data does not matter, and *last* here means the last observation chronologically.

intervals(w | *numlist*) specifies the intervals into which the data are to be aggregated for tabular presentation. A numeric argument is interpreted as the width of the interval. For instance, interval(2) aggregates data into the intervals $0 \le t < 2$, $2 \le t < 4$, and so on. Not specifying interval() is equivalent to specifying interval(1). Because in most data, failure times are

recorded as integers, this amounts to no aggregation except that implied by the recording of the time variable, and so it produces Kaplan–Meier product-limit estimates of the survival curve (with an actuarial adjustment; see the noadjust option above). Also see [ST] sts list. Although it is possible to examine survival and failure without aggregation, some form of aggregation is almost always required to examine the hazard.

When more than one argument is specified, intervals are aggregated as specified. For instance, interval(0,2,8,16) aggregates data into the intervals $0 \le t < 2$, $2 \le t < 8$, and $8 \le t < 16$, and (if necessary) the open-ended interval $t \ge 16$.

interval(w) is equivalent to interval(0,7,15,30,60,90,180,360,540,720), corresponding to 1 week, (roughly) 2 weeks, 1 month, 2 months, 3 months, 6 months, 1 year, 1.5 years, and 2 years when failure times are recorded in days. The w suggests widening intervals.

saving(*filename* [, replace]) creates a Stata data file (.dta file) containing the life table. This option will not save the graph to disk; see [G-2] graph save to save the resulting graph to disk.

  replace specifies that *filename* be overwritten if it exists. This option is not shown in the dialog box.

Plot

plotopts(*plot_options*) affects the rendition of the plotted line and plotted points; see [G-3] connect_options and [G-3] marker_options.

plot#opts(*plot_options*) affects the rendition of the #th plotted line and plotted points; see [G-3] connect_options and [G-3] marker_options. This option is valid only if overlay is specified.

CI plot

ciopts(*rspike_options*) affects the rendition of the confidence intervals for the graphed survival, failure, or hazard; see [G-3] rspike_options.

ci#opts(*rspike_options*) affects the rendition of the #th confidence interval for the graphed survival, failure, or hazard; see [G-3] rspike_options. This option is valid only if overlay is specified.

Add plots

addplot(*plot*) provides a way to add other plots to the generated graph; see [G-3] addplot_option.

Y axis, X axis, Titles, Legend, and Overall

twoway_options are any of the options documented in [G-3] twoway_options, excluding by(). These include options for titling the graph (see [G-3] title_options) and for saving the graph to disk (see [G-3] saving_option).

byopts(*byopts*) affects the appearance of the combined graph when by() is specified, including the overall graph title and the organization of subgraphs. See [G-3] by_option.

Remarks and examples   stata.com

Life tables describe death rates in a given population over time. Such tables date back to the 17th century; Edmund Halley (1693) is often credited with their development. ltable is for use with "cohort" data, and although one often thinks of such tables as monitoring a population from the "birth" of the first member to the "death" of the last, more generally, such tables can be thought of as a reasonable way to list any kind of survival data. For an introductory discussion of life tables, see Pagano and Gauvreau (2000, 489–495) and Oliveira (2013); for an intermediate discussion, see Selvin (2004, 335–377); and for a more complete discussion, see Chiang (1984).

▷ Example 1

In Pike (1966), two groups of rats were exposed to a carcinogen, and the number of days to death from vaginal cancer was recorded (reprinted in Kalbfleisch and Prentice 2002, 2):

Group 1	143	164	188	188	190	192	206	209	213	216
	220	227	230	234	246	265	304	216*	244*	
Group 2	142	156	163	198	205	232	232	233	233	233
	233	239	240	261	280	280	296	296	323	204*
	344*									

The '*' on a few of the entries indicates that the observation was censored—as of the recorded day, the rat had still not died because of vaginal cancer but was withdrawn from the experiment for other reasons.

Having entered these data into Stata, we see that the first few observations are

```
. use http://www.stata-press.com/data/r14/rat
. list in 1/5
```

	group	t	died
1.	1	143	1
2.	1	164	1
3.	1	188	1
4.	1	188	1
5.	1	190	1

For example, the first observation records a rat from group 1 that died on the 143rd day. The died variable records whether that rat died or was withdrawn (censored):

```
. list if died==0
```

	group	t	died
18.	1	216	0
19.	1	244	0
39.	2	204	0
40.	2	344	0

Four rats, two from each group, did not die but were withdrawn.

The life table for group 1 is

```
. ltable t died if group==1
                         Beg.                             Std.
    Interval            Total   Deaths   Lost   Survival   Error    [95% Conf. Int.]

    143    144            19       1      0     0.9474    0.0512    0.6812    0.9924
    164    165            18       1      0     0.8947    0.0704    0.6408    0.9726
    188    189            17       2      0     0.7895    0.0935    0.5319    0.9153
    190    191            15       1      0     0.7368    0.1010    0.4789    0.8810
    192    193            14       1      0     0.6842    0.1066    0.4279    0.8439
    206    207            13       1      0     0.6316    0.1107    0.3790    0.8044
    209    210            12       1      0     0.5789    0.1133    0.3321    0.7626
    213    214            11       1      0     0.5263    0.1145    0.2872    0.7188
    216    217            10       1      1     0.4709    0.1151    0.2410    0.6713
    220    221             8       1      0     0.4120    0.1148    0.1937    0.6194
    227    228             7       1      0     0.3532    0.1125    0.1502    0.5648
    230    231             6       1      0     0.2943    0.1080    0.1105    0.5070
    234    235             5       1      0     0.2355    0.1012    0.0751    0.4459
    244    245             4       0      1     0.2355    0.1012    0.0751    0.4459
    246    247             3       1      0     0.1570    0.0931    0.0312    0.3721
    265    266             2       1      0     0.0785    0.0724    0.0056    0.2864
    304    305             1       1      0     0.0000      .          .         .
```

The reported survival rates are the survival rates at the end of the interval. Thus, 94.7% of rats survived 144 days or more.

◁

❑ Technical note

If you compare the table just printed with the corresponding table in Kalbfleisch and Prentice (2002, 16), you will notice that the survival estimates differ beginning with the interval 216–217, which is the first interval containing a censored observation. ltable treats censored observations as if they were withdrawn halfway through the interval. The table printed in Kalbfleisch and Prentice treated censored observations as if they were withdrawn at the end of the interval, even though Kalbfleisch and Prentice (2002, 19) mention how results could be adjusted for censoring.

Here the same results as those printed in Kalbfleisch and Prentice could be obtained by incrementing the time of withdrawal by 1 for the four censored observations. We say "here" because there were no deaths on the incremented dates. For instance, one of the rats was withdrawn on the 216th day, a day on which there was also a real death. There were no deaths on day 217, however, so moving the withdrawal forward 1 day is equivalent to assuming that the withdrawal occurred at the end of the day 216–217 interval. If the adjustments are made and ltable is used to calculate survival in both groups, the results are the same as those printed in Kalbfleisch and Prentice, except that for group 2 in the interval 240–241, they report the survival as 0.345 when they mean 0.354.

In any case, the one-half adjustment for withdrawals is generally accepted, but it is only a crude adjustment that becomes cruder the wider the intervals.

❑

▷ Example 2: ltable with aggregated intervals

When you do not specify the intervals, ltable uses unit intervals. The only aggregation performed on the data was aggregation due to deaths or withdrawals occurring on the same "day". If we wanted to see the table aggregated into 30-day intervals, we would type

```
. ltable t died if group==1, interval(30)
```

Interval		Beg. Total	Deaths	Lost	Survival	Std. Error	[95% Conf. Int.]	
120	150	19	1	0	0.9474	0.0512	0.6812	0.9924
150	180	18	1	0	0.8947	0.0704	0.6408	0.9726
180	210	17	6	0	0.5789	0.1133	0.3321	0.7626
210	240	11	6	1	0.2481	0.1009	0.0847	0.4552
240	270	4	2	1	0.1063	0.0786	0.0139	0.3090
300	330	1	1	0	0.0000	.	.	.

The interval displayed as 120 150 indicates the interval including 120 and up to, but not including, 150. The reported survival rate is the survival rate just after the close of the interval.

When you specify more than one number as the argument to interval(), you specify the cutoff points, not the widths.

```
. ltable t died if group==1, interval(120,180,210,240,330)
```

Interval		Beg. Total	Deaths	Lost	Survival	Std. Error	[95% Conf. Int.]	
120	180	19	2	0	0.8947	0.0704	0.6408	0.9726
180	210	17	6	0	0.5789	0.1133	0.3321	0.7626
210	240	11	6	1	0.2481	0.1009	0.0847	0.4552
240	330	4	3	1	0.0354	0.0486	0.0006	0.2245

If any of the underlying failure or censoring times are larger than the last cutoff specified, then they are treated as being in the open-ended interval:

```
. ltable t died if group==1, interval(120,180,210,240)
```

Interval		Beg. Total	Deaths	Lost	Survival	Std. Error	[95% Conf. Int.]	
120	180	19	2	0	0.8947	0.0704	0.6408	0.9726
180	210	17	6	0	0.5789	0.1133	0.3321	0.7626
210	240	11	6	1	0.2481	0.1009	0.0847	0.4552
240	.	4	3	1	0.0354	0.0486	0.0006	0.2245

Whether the last interval is treated as open ended or not makes no difference for survival and failure tables, but it does affect hazard tables. If the interval is open ended, the hazard is not calculated for it.

◁

▷ Example 3: ltable with separate tables for each group

The by(*varname*) option specifies that separate tables be presented for each value of *varname*. Remember that our rat dataset contains two groups:

```
. ltable t died, by(group) interval(30)
```

Interval		Beg. Total	Deaths	Lost	Survival	Std. Error	[95% Conf. Int.]	
group = 1								
120	150	19	1	0	0.9474	0.0512	0.6812	0.9924
150	180	18	1	0	0.8947	0.0704	0.6408	0.9726
180	210	17	6	0	0.5789	0.1133	0.3321	0.7626
210	240	11	6	1	0.2481	0.1009	0.0847	0.4552
240	270	4	2	1	0.1063	0.0786	0.0139	0.3090
300	330	1	1	0	0.0000	.	.	.
group = 2								
120	150	21	1	0	0.9524	0.0465	0.7072	0.9932
150	180	20	2	0	0.8571	0.0764	0.6197	0.9516
180	210	18	2	1	0.7592	0.0939	0.5146	0.8920
210	240	15	7	0	0.4049	0.1099	0.1963	0.6053
240	270	8	2	0	0.3037	0.1031	0.1245	0.5057
270	300	6	4	0	0.1012	0.0678	0.0172	0.2749
300	330	2	1	0	0.0506	0.0493	0.0035	0.2073
330	360	1	0	1	0.0506	0.0493	0.0035	0.2073

◁

▷ Example 4: ltable for failure tables

A failure table is simply a different way of looking at a survival table; failure is 1 − survival:

```
. ltable t died if group==1, interval(30) failure
```

Interval		Beg. Total	Deaths	Lost	Cum. Failure	Std. Error	[95% Conf. Int.]	
120	150	19	1	0	0.0526	0.0512	0.0076	0.3188
150	180	18	1	0	0.1053	0.0704	0.0274	0.3592
180	210	17	6	0	0.4211	0.1133	0.2374	0.6679
210	240	11	6	1	0.7519	0.1009	0.5448	0.9153
240	270	4	2	1	0.8937	0.0786	0.6910	0.9861
300	330	1	1	0	1.0000	.	.	.

◁

▷ Example 5: Survival rate at start of interval versus end of interval

Selvin (2004, 357) presents follow-up data from Cutler and Ederer (1958) on six cohorts of kidney cancer patients. The goal is to estimate the 5-year survival probability.

Year	Interval	Alive	Deaths	Lost	With-drawn	Year	Interval	Alive	Deaths	Lost	With-drawn
1946	0−1	9	4	1		1948	0−1	21	11	0	
	1−2	4	0	0			1−2	10	1	2	
	2−3	4	0	0			2−3	7	0	0	
	3−4	4	0	0			3−4	7	0	0	7
	4−5	4	0	0		1949	0−1	34	12	0	
	5−6	4	0	0	4		1−2	22	3	3	
1947	0−1	18	7	0			2−3	16	1	0	15
	1−2	11	0	0		1950	0−1	19	5	1	
	2−3	11	1	0			1−2	13	1	1	11
	3−4	10	2	2		1951	0−1	25	8	2	15
	4−5	6	0	0	6						

The following is the Stata dataset corresponding to the table:

```
. use http://www.stata-press.com/data/r14/selvin

. list
```

	year	t	died	pop
1.	1946	.5	1	4
2.	1946	.5	0	1
3.	1946	5.5	0	4
4.	1947	.5	1	7
5.	1947	2.5	1	1
	(output omitted)			

As summary data may often come in the form shown above, it is worth understanding exactly how the data were translated for use with `ltable`. `t` records the time of death or censoring (lost to follow-up or withdrawal). `died` contains 1 if the observation records a death and 0 if it instead records lost or withdrawn patients. `pop` records the number of patients in the category. The first line of the original table stated that, in the 1946 cohort, there were nine patients at the start of the interval 0−1, and during the interval, four died and one was lost to follow-up. Thus we entered in observation 1 that at `t` = 0.5, four patients died and in observation 2 that at `t` = 0.5, one patient was censored. We ignored the information on the total population because `ltable` will figure that out for itself.

The second line of the table indicated that in the interval 1−2, four patients were still alive at the beginning of the interval, and during the interval, zero died or were lost to follow-up. Because no patients died or were censored, we entered nothing into our data. Similarly, we entered nothing for lines 3, 4, and 5 of the table. The last line for 1946 stated that, in the interval 5−6, four patients were alive at the beginning of the interval and that those four patients were withdrawn. In observation 3, we entered that there were four censorings at `t` = 5.5.

It does not matter that we chose to record the times of deaths or censoring as midpoints of intervals; we could just as well have recorded the times as 0.8 and 5.8. By default, `ltable` will form intervals 0−1, 1−2, and so on, and place observations into the intervals to which they belong. We suggest using 0.5 and 5.5 because those numbers correspond to the underlying assumptions made by `ltable` in making its calculations. Using midpoints reminds you of these assumptions.

To obtain the survival rates, we type

```
. ltable t died [freq=pop]
```

	Interval	Beg. Total	Deaths	Lost	Survival	Std. Error	[95% Conf. Int.]	
0	1	126	47	19	0.5966	0.0455	0.5017	0.6792
1	2	60	5	17	0.5386	0.0479	0.4405	0.6269
2	3	38	2	15	0.5033	0.0508	0.4002	0.5977
3	4	21	2	9	0.4423	0.0602	0.3225	0.5554
4	5	10	0	6	0.4423	0.0602	0.3225	0.5554
5	6	4	0	4	0.4423	0.0602	0.3225	0.5554

We estimate the 5-year survival rate as 0.4423 and the 95% confidence interval as 0.3225 to 0.5554.

Selvin (2004, 361), in presenting these results, lists the survival in the interval 0−1 as 1, in 1−2 as 0.597, in 2−3 as 0.539, and so on. That is, relative to us, he shifted the rates down one row and inserted a 1 in the first row. In his table, the survival rate is the survival rate at the *start* of the interval. In our table, the survival rate is the survival rate at the *end* of the interval (or, equivalently, at the start of the next interval). This is, of course, simply a difference in the way the numbers are presented and not in the numbers themselves.   ◁

▷ Example 6: ltable for hazard tables

The discrete hazard function is the rate of failure—the number of failures occurring within a time interval divided by the width of the interval (assuming that there are no censored observations). Although the survival and failure tables are meaningful at the "individual" level—with intervals so narrow that each contains only one failure—that is not true for the discrete hazard. If all intervals contained one death and if all intervals were of equal width, the hazard function would be $1/\Delta t$ and so appear to be a constant!

The empirically determined discrete hazard function can be revealed only by aggregation. Gross and Clark (1975, 37) print data on malignant melanoma at the University of Texas M. D. Anderson Tumor Clinic between 1944 and 1960. The interval is the time from initial diagnosis:

Interval (years)	Number lost to follow-up	Number with-drawn alive	Number dying
0−1	19	77	312
1−2	3	71	96
2−3	4	58	45
3−4	3	27	29
4−5	5	35	7
5−6	1	36	9
6−7	0	17	3
7−8	2	10	1
8−9	0	8	3
9+	0	0	32

For our statistical purposes, there is no difference between the number lost to follow-up (patients who disappeared) and the number withdrawn alive (patients dropped by the researchers)—both are censored. We have entered the data into Stata; here are a few of the data:

```
. use http://www.stata-press.com/data/r14/tumor

. list in 1/6, separator(0)
```

	t	d	pop
1.	.5	1	312
2.	.5	0	19
3.	.5	0	77
4.	1.5	1	96
5.	1.5	0	3
6.	1.5	0	71

We entered each group's time of death or censoring as the midpoint of the intervals and entered the numbers of the table, recording d as 1 for deaths and 0 for censoring. The hazard table is
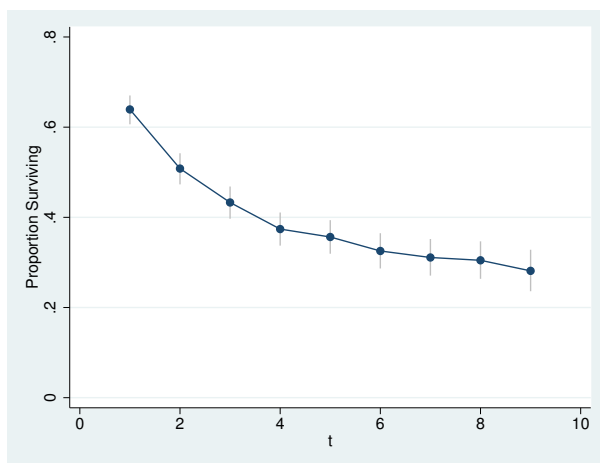
```
. ltable t d [freq=pop], hazard interval(0(1)9)
```

Interval		Beg. Total	Cum. Failure	Std. Error	Hazard	Std. Error	[95% Conf. Int.]	
0	1	913	0.3607	0.0163	0.4401	0.0243	0.3924	0.4877
1	2	505	0.4918	0.0176	0.2286	0.0232	0.1831	0.2740
2	3	335	0.5671	0.0182	0.1599	0.0238	0.1133	0.2064
3	4	228	0.6260	0.0188	0.1461	0.0271	0.0931	0.1991
4	5	169	0.6436	0.0190	0.0481	0.0182	0.0125	0.0837
5	6	122	0.6746	0.0200	0.0909	0.0303	0.0316	0.1502
6	7	76	0.6890	0.0208	0.0455	0.0262	0.0000	0.0969
7	8	56	0.6952	0.0213	0.0202	0.0202	0.0000	0.0598
8	9	43	0.7187	0.0235	0.0800	0.0462	0.0000	0.1705
9	.	32	1.0000	.	.	.	.	.

We specified the interval() option as we did—and not as interval(1) or omitting the option altogether—to force the last interval to be open ended. Had we not, and if we had recorded t as 9.5 for observations in that interval (as we did), ltable would have calculated a hazard rate for the "interval". Here the result of that calculation would have been 2, but no matter the result, it would have been meaningless because we do not know the width of the interval.

When dealing with the survivor or failure function, you are not limited to merely examining a column of numbers. With the graph option, you can see the result graphically:

```
. ltable t d [freq=pop], i(0(1)9) graph notable ci xlab(0(2)10)
```


The vertical lines in the graph represent the 95% confidence intervals for the survivor function. Among the options we specified, although it is not required, is `notable`, which suppressed printing the table, saving us some paper. `xlab()` was passed through to the `graph` command (see [G-3] *twoway_options*) and was unnecessary but made the graph look better.

◁

❑ Technical note

Because many intervals can exist during which no failures occur (in which case the hazard estimate is zero), the estimated hazard is best graphically represented using a kernel smooth. Such an estimate is available in `sts graph`; see [ST] **sts graph**.

❑

Video example

How to construct life tables

Methods and formulas

Let τ_i be the individual failure or censoring times. The data are aggregated into intervals given by t_j, $j = 1, \ldots, J$, and $t_{J+1} = \infty$ with each interval containing counts for $t_j \le \tau < t_{j+1}$. Let d_j be the number of failures during the interval, m_j be the censored observations during the interval, and N_j be the number alive at the start of the interval. Define $n_j = N_j - m_j/2$ as the adjusted number at risk at the start of the interval. If the `noadjust` option is specified, $n_j = N_j$.

The product-limit estimate of the survivor function is

$$S_j = \prod_{k=1}^{j} \frac{n_k - d_k}{n_k}$$

(Kalbfleisch and Prentice 2002, 10, 15). Greenwood's formula for the asymptotic standard error of S_j is

$$s_j = S_j \sqrt{\sum_{k=1}^{j} \frac{d_k}{n_k(n_k - d_k)}}$$

(Greenwood 1926; Kalbfleisch and Prentice 2002, 17). s_j is reported as the standard deviation of survival but is not used in generating the confidence intervals because it can produce intervals outside 0 and 1. The "natural" units for the survivor function are $\log(-\log S_j)$, and the asymptotic standard error of that quantity is

$$\widehat{s}_j = \sqrt{\frac{\sum d_k / \{n_k(n_k - d_k)\}}{\left[\sum \log\left\{(n_k - d_k)/n_k\right\}\right]^2}}$$

(Kalbfleisch and Prentice 2002, 18). The corresponding confidence intervals are $S_j^{\exp(\pm z_{1-\alpha/2}\widehat{s}_j)}$.

The cumulative failure time is defined as $G_j = 1 - S_j$, and thus the variance is the same as for S_j and the confidence intervals are $1 - S_j^{\exp(\pm z_{1-\alpha/2}\widehat{s}_j)}$.

Both S_j and G_j are graphed against t_{j+1}.

Define the within-interval failure rate as $f_j = d_j/n_j$. The maximum likelihood estimate of the (within-interval) hazard is then

$$\lambda_j = \frac{f_j}{(1 - f_j/2)(t_{j+1} - t_j)}$$

The standard error of λ_j is

$$s_{\lambda_j} = \lambda_j \sqrt{\frac{1 - \{(t_{j+1} - t_j)\lambda_j/2\}^2}{d_j}}$$

from which a confidence interval is calculated.

If the noadjust option is specified, the estimate of the hazard is

$$\lambda_j = \frac{f_j}{t_{j+1} - t_j}$$

and its standard error is

$$s_{\lambda_j} = \frac{\lambda_j}{\sqrt{d_j}}$$

The confidence interval is

$$\left[\frac{\lambda_j}{2d_j}\chi^2_{2d_j,\alpha/2} \,,\; \frac{\lambda_j}{2d_j}\chi^2_{2d_j,1-\alpha/2}\right]$$

where $\chi^2_{2d_j,q}$ is the qth quantile of the χ^2 distribution with $2d_j$ degrees of freedom (Cox and Oakes 1984, 53–54, 38–40).

For the likelihood-ratio test for homogeneity, let d_g be the total number of deaths in the gth group. Define $T_g = \sum_{i \in g} \tau_i$, where i indexes the individual failure or censoring times. The χ^2 value with $G - 1$ degrees of freedom (where G is the total number of groups) is

$$\chi^2 = 2 \left\{ \left(\sum d_g \right) \log \left(\frac{\sum T_g}{\sum d_g} \right) - \sum d_g \log \left(\frac{T_g}{d_g} \right) \right\}$$

(Lawless 2003, 155).

The log-rank test for homogeneity is the test presented by `sts test`; see [ST] **sts**.

Acknowledgments

References

Chiang, C. L. 1984. *The Life Table and Its Applications*. Malabar, FL: Krieger.

Cox, D. R., and D. Oakes. 1984. *Analysis of Survival Data*. London: Chapman & Hall/CRC.

Cutler, S. J., and F. Ederer. 1958. Maximum utilization of the life table method in analyzing survival. *Journal of Chronic Diseases* 8: 699–712.

Greenwood, M. 1926. The natural duration of cancer. *Reports on Public Health and Medical Subjects* 33: 1–26.

Gross, A. J., and V. A. Clark. 1975. *Survival Distribution: Reliability Applications in the Biomedical Sciences*. New York: Wiley.

Halley, E. 1693. An estimate of the degrees of the mortality of mankind, drawn from curious tables of the births and funerals at the city of Breslaw; with an attempt to ascertain the price of annuities upon lives. *Philosophical Transactions* 17: 596–610.

Kahn, H. A., and C. T. Sempos. 1989. *Statistical Methods in Epidemiology*. New York: Oxford University Press.

Kalbfleisch, J. D., and R. L. Prentice. 2002. *The Statistical Analysis of Failure Time Data*. 2nd ed. New York: Wiley.

Krakauer, H., and J. Stewart. 1991. ssa1: Actuarial or life-table analysis of time-to-event data. *Stata Technical Bulletin* 1: 23–25. Reprinted in *Stata Technical Bulletin Reprints*, vol. 1, pp. 200–202. College Station, TX: Stata Press.

Lawless, J. F. 2003. *Statistical Models and Methods for Lifetime Data*. 2nd ed. New York: Wiley.

Oliveira, A. G. 2013. *Biostatistics Decoded*. Chichester, UK: Wiley.

Pagano, M., and K. Gauvreau. 2000. *Principles of Biostatistics*. 2nd ed. Belmont, CA: Duxbury.

Pike, M. C. 1966. A method of analysis of a certain class of experiments in carcinogenesis. *Biometrics* 22: 142–161.

Ramalheira, C. 2001. ssa14: Global and multiple causes-of-death life tables. *Stata Technical Bulletin* 59: 29–45. Reprinted in *Stata Technical Bulletin Reprints*, vol. 10, pp. 333–355. College Station, TX: Stata Press.

Selvin, S. 2004. *Statistical Analysis of Epidemiologic Data*. 3rd ed. New York: Oxford University Press.

Also see

[ST] **stcox** — Cox proportional hazards model