How to Build the Kaplan-Meier Curve from the Ground Up

Presented by David Franklin, 30-Aug-06

The Kaplan-Meier Survival Curve is used to show the time until subjects in a study present a specific event or endpoint. This event can be death, appearance of an adverse event, etcetera.

The Kaplan-Meier estimator is defined as

$$\hat{S}(t) = \prod_{ti \le t} \left(1 - \frac{di}{ni}\right)$$

where *ti* is the duration of study at point *i*, *di* is the number of deaths up to point *i*, and *ni* is the number of individuals at risk just prior to *ti*. S is based on the probability that an individual survives at the end of a time interval, on the condition that the individual was present at the start of the time interval.

Two terms are used to classify a patient when looking at this data:

event	the "event" occurred during the study at a particular timepoint for a subject
censor	mathematically removing a subject from the curve at the end of their time on the study if the subject did not have an event

Example
The survival times for ten subjects is listed below in ascending order:

2, 15+, 17, 18, 18+, 20+, 23, 25+, 30+, 31

The '+' sign signifies that the patient was alive at the end of the study and after any follow-up.

Day	Censor	Number at Risk at Start of Time	Died During Interval	Proportion Surviving Interval	Cumulative Survival at End of Interval
0		10	0	1.0000	1.0000
2	1	10	1	0.9000	0.9000
15	0				
17	1	8	1	0.8750	0.9000*0.8750=0.7875
18	1				
18	0	7	1	0.8571	0.7875*0.8571=0.6750
20	0				
23	1	4	1	0.7500	0.6750*0.7500=0.5063
25	0				
30	0				
31	1	1	1	0.0000	0.5063*0.0000=0.0000

Note that the calculations are done for days where subjects actually died - at other days results are 'censored' and the survival calculations do not change. Also note that in the example the censor variable is set to '1' if the subject died and '0' if the subject was alive at the end of the period we are looking at.

Results from the Cumulative Survival column plot the Kaplan-Meier Curve.

Doing the Kaplan-Meier Curve in SAS is done using the LIFETEST procedure. Using the same data as above the following SAS code was used:

```
title1 "Kaplan-Meier Curve - Time to Death";
proc lifetest data=can0 plots=(s);
    time time*censor(0);
run;
```

resulting in the following output:

The LIFETEST Procedure
Product-Limit Survival Estimates

time	Survival	Failure	Survival Standard Error	Number Failed	Number Left
0.0000	1.0000	0	0	0	10
2.0000	0.9000	0.1000	0.0949	1	9
15.0000*	.	.	.	1	8
17.0000	0.7875	0.2125	0.1340	2	7
18.0000	0.6750	0.3250	0.1551	3	6
18.0000*	.	.	.	3	5
20.0000*	.	.	.	3	4
23.0000	0.5063	0.4938	0.1868	4	3
25.0000*	.	.	.	4	2
30.0000*	.	.	.	4	1
31.0000	0	1.0000	0	5	0

NOTE: The marked survival times are censored observations.

Summary Statistics for Time Variable time

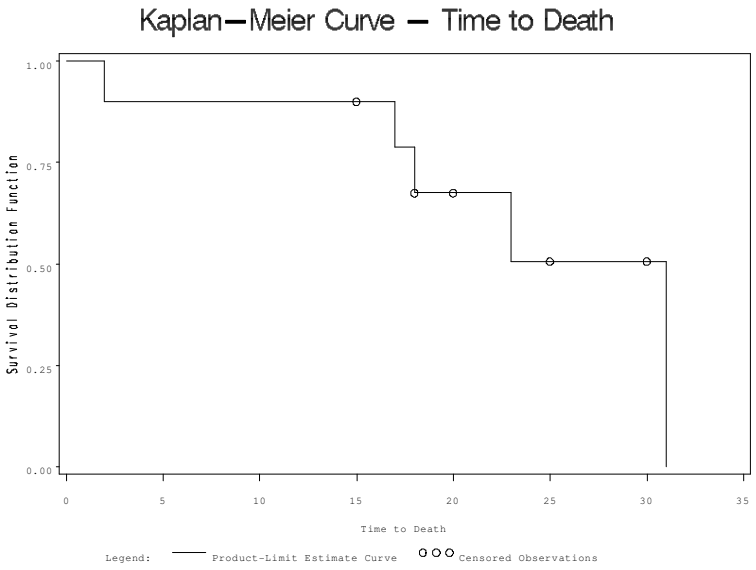
Quartile Estimates

Percent	Point Estimate	95% Confidence Interval [Lower	Upper)
75	31.0000	23.0000	31.0000
50	31.0000	18.0000	31.0000
25	18.0000	2.0000	31.0000

Mean	Standard Error
23.7125	3.3768

Summary of the Number of Censored and Uncensored Values

Total	Failed	Censored	Percent Censored
10	5	5	50.00



The plot from the LIFETEST procedure can be enhanced by using the ANNOTATE= option in the LIFETEST statement.

If two or more treatment groups are to be plotted on the one plot, the STRATA statement in the LIFETEST procedure can be used, for example if there were a treatment variable (variable name TRTMNT) in the data above the SAS code would be:

```
proc lifetest data=can0 plots=(s);
    time time*censor(0);
    strata trtmnt;
run;
```

/*End of File*/

The Cumulative Hazard Curve

The Cumulative Hazard estimate is calculated as

$$\hat{H}(t) = -LN(\hat{S}(t))$$

To plot the curve just replace the

```
plots=(s)
```

option with

```
plots=(h)
```

in the LIFETEST procedure call. Using the example in this paper the code would be:

```
proc lifetest
      data=can0
      plots=(h);
 time time*censor(0);
run;
```

SAS Tip
Format a date to dd-mmm-yyyy

```
40 proc format;
41   picture DatePt
42     low-high='%d-%b-%YYYY'
43         (datatype=date);
NOTE: Format DATEPT has
been output.
44   run;
NOTE: PROCEDURE FORMAT
used:
      real time          0.01 seconds
      cpu time           0.01 seconds

45   data _null_;
46       today="&sysdate9"d;
47       put today DatePt.;
48   run;
30-AUG-2006
NOTE: DATA statement used:
      real time          0.00 seconds
      cpu time           0.00 seconds
```

SAS Trivia

Development for SAS started in 1966 after a grant from NIH to develop general purpose statistical software. The original developers were Jim Goodnight, John Sal, Jane Helwig and Tony Barr. The first release was in 1972 for the IBM mainframe at NC State University. The original system was written in Fortran. Development of SAS continued with USDA funding until the SAS Institute was founded in 1976 by current CEO and President Jim Goodnight.