Purpose:
Estimating survival curves with PROC LIFETEST in SAS

About 30 years ago, before Cox (1972) proposed Proportional Hazard Regression model, estimating survival curves was the dominant method of survival analysis. Nowadays, estimation of survival curves is still very useful in exploratory data analysis.

The Kaplan-Meier method is the most famous technique for estimating survival functions. Note that if our failure time data is complete, i.e., no censoring in the data set, the K-M estimate is reduced to empirical estimate of survival function. In this lab we'll talk about how to use SAS to derive the estimation of survival function and graph it.

1. Estimating survival curves with PROC LIFETEST in SAS

We use the same data set, "**myelomatosis data**", which was given in the first lab to illustrate how to get the K-M estimator using PROC LIFETEST.

```
DATA myel;
      SET "C:\courses\Survival\2006\labs\lab2\myel";
RUN;

PROC LIFETEST data=myel;
      TIME dur*status(0);
RUN;
```

The K-M estimator is the default, so you do not need to request it. To be explicit, you can put METHOD=KM in the PROC LIFETEST statement. In TIME statement, the first variable is the time of the event or censoring; the second variable contains information on whether or not the observation was censored; and the number in parentheses are values of the second variable corresponding to censored observations. Specifically, in this data, "dur" is the time of the event or censoring; "status" is the indicator containing information on whether the observation was censored or not, with 0 corresponding to censoring. The result of the previous command is as follows:

The LIFETEST Procedure

Product-Limit Survival Estimates

dur	Survival	Failure	Survival Standard Error	Number Failed	Number Left
0.00	1.0000	0	0	0	25
8.00	.	.	.	1	24
8.00	0.9200	0.0800	0.0543	2	23
13.00	0.8800	0.1200	0.0650	3	22
18.00	0.8400	0.1600	0.0733	4	21
23.00	0.8000	0.2000	0.0800	5	20
52.00	0.7600	0.2400	0.0854	6	19
63.00	.	.	.	7	18
63.00	0.6800	0.3200	0.0933	8	17
70.00	0.6400	0.3600	0.0960	9	16
76.00	0.6000	0.4000	0.0980	10	15
180.00	0.5600	0.4400	0.0993	11	14
195.00	0.5200	0.4800	0.0999	12	13

210.00	0.4800	0.5200	0.0999	13	12
220.00	0.4400	0.5600	0.0993	14	11
365.00*	.	.	.	14	10
632.00	0.3960	0.6040	0.0986	15	9
700.00	0.3520	0.6480	0.0970	16	8
852.00*	.	.	.	16	7
1296.00	0.3017	0.6983	0.0953	17	6
1296.00*	.	.	.	17	5
1328.00*	.	.	.	17	4
1460.00*	.	.	.	17	3
1976.00*	.	.	.	17	2
1990.00*	.	.	.	17	1
2240.00*	.	.	.	17	0

NOTE: The marked survival times are censored observations.

Summary Statistics for Time Variable dur

Quartile Estimates

Percent	Point Estimate	95% Confidence Interval [Lower	Upper)
75	.	220.00	.
50	210.00	63.00	1296.00
25	63.00	18.00	195.00

Mean	Standard Error
562.76	117.32

NOTE: The mean survival time and its standard error were underestimated because the largest observation was censored and the estimation was restricted to the largest event time.

Summary of the Number of Censored and Uncensored Values

Total	Failed	Censored	Percent Censored
25	17	8	32.00

- Each line of numbers corresponds to an observation (except for the first row, which is for time 0), arranged in ascending order. Censored observations are stared.
- The K-M estimates are given in the second column, e.g., at 195 days, the K-M estimate of survival probability is 0.52 with standard error 0.0999. When there are ties (at 8 days and 63 days), the K-M estimate is reported only for the last of the tied cases. No K-M estimates are reported for the censored times.
- K-M estimator is defined for any time between 0 and the largest event or censoring time. It's just that it only changes at an observed event time. Thus, the estimated survival probability for any time from 195 days and 210 days is 0.52.
- If the largest observation is a censored case, the K-M estimate after that time point is undefined. In this case, there is no K-M estimate for the time after 2240 days.
- The third column, named FAILED, is just 1-SURVIVAL, which is the estimated probability of a death prior to the specified time.

- The forth column, labeled SURVIVAL STANDARD ERROR, is an estimate of the standard error of the K-M estimate, obtained by Greenwood's formula.
- The fifth column is the cumulative number of cases that experienced events prior to and including the specific time point. The last column is the number of cases that have neither experienced events nor been censored prior to that time point.
- The Quartile Estimates in the next section give the estimated 75th, 50th and 25th percentiles. You can easily get these estimates from the previous table by yourself. For this data, no value is reported for the 75th percentile because the K-M estimator for these data never reaches a failure probability greater than 0.70. The 50th percentile is also called median death time, which is often of greatest interest in the study.
- An estimated mean time of death is also reported. Note that the median estimate is preferable than the mean estimate since when a large proportion of the cases are censored, the upper tail of the distribution will be poorly estimated.

To get confidence intervals around the survival probabilities, you can calculate them by hand using the standard errors reported in the table. However, it's much easier to write the intervals to an output data set by using the OUTSURV=<generic name of the output file> option in the PROC LIFETEST statement and print them out:

```
PROC LIFETEST data=myel outsurv=myoutput;
      TIME dur*status(0);
RUN;

PROC print data=myoutput;
RUN;
```

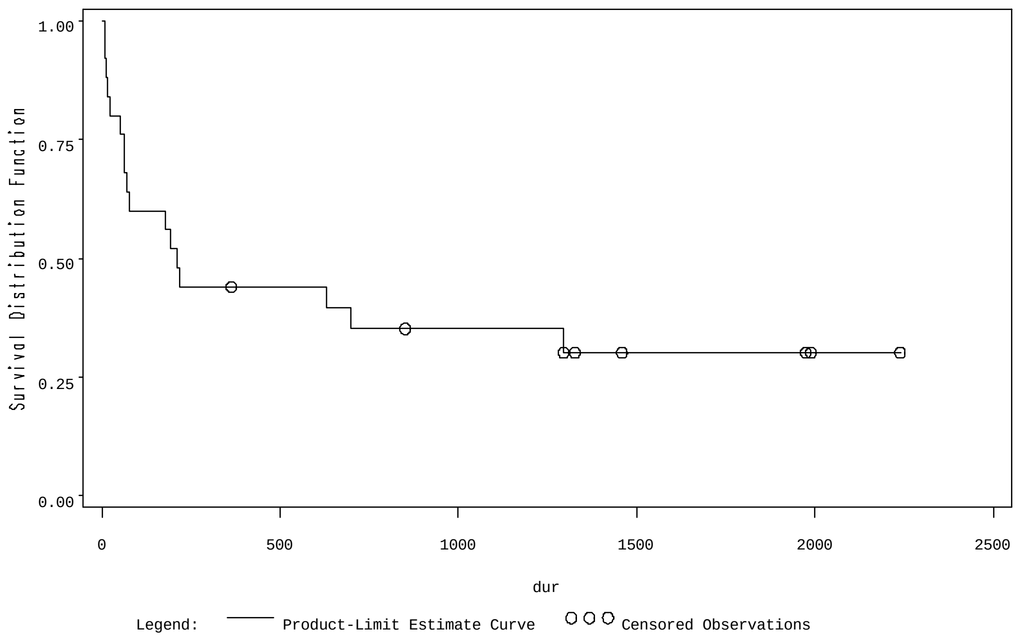
Obs	dur	_CENSOR_	SURVIVAL	SDF_LCL	SDF_UCL
1	0	0	1.00000	1.00000	1.00000
2	8	0	0.92000	0.81366	1.00000
3	13	0	0.88000	0.75262	1.00000
4	18	0	0.84000	0.69629	0.98371
5	23	0	0.80000	0.64320	0.95680
6	52	0	0.76000	0.59259	0.92741
7	63	0	0.68000	0.49715	0.86285
8	70	0	0.64000	0.45184	0.82816
9	76	0	0.60000	0.40796	0.79204
10	180	0	0.56000	0.36542	0.75458
11	195	0	0.52000	0.32416	0.71584
12	210	0	0.48000	0.28416	0.67584
13	220	0	0.44000	0.24542	0.63458
14	365	1	0.44000	.	.
15	632	0	0.39600	0.20271	0.58929
16	700	0	0.35200	0.16192	0.54208
17	852	1	0.35200	.	.
18	1296	0	0.30171	0.11498	0.48845
19	1296	1	.	.	.
20	1328	1	.	.	.
21	1460	1	.	.	.
22	1976	1	.	.	.
23	1990	1	.	.	.
24	2240	1	.	.	.

The last two columns are the lower and upper limits of the 95% confidence interval.

You can get a plot of the estimated survival function by requesting it in the PROC LIFETEST statement:

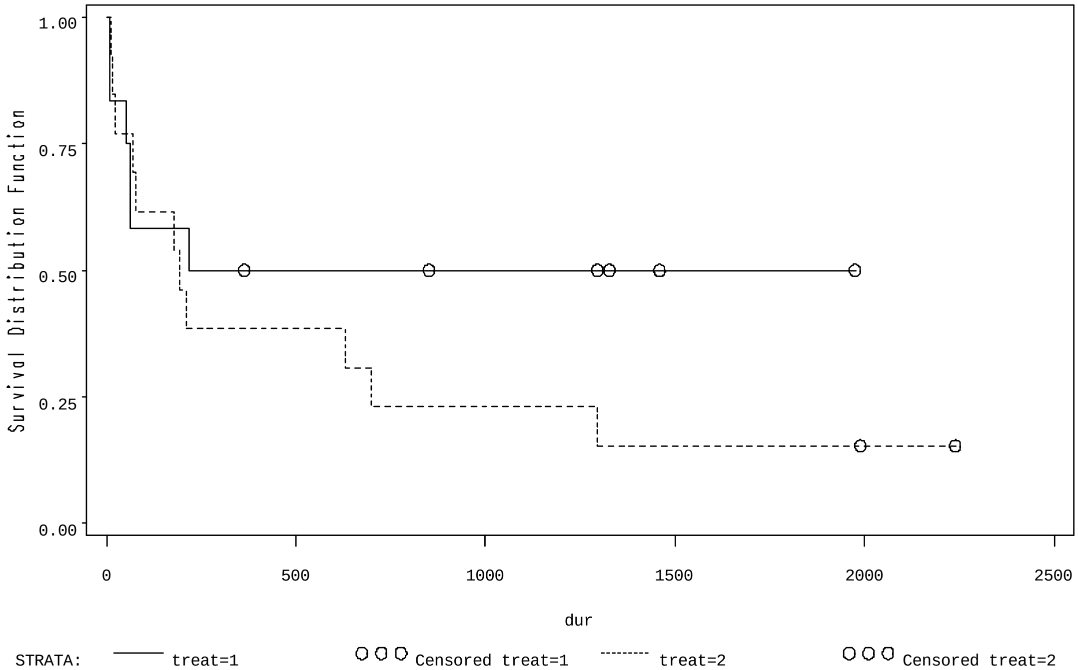
```
PROC LIFETEST data=myel plots=(s) graphics;
      TIME dur*status(0);
RUN;
```

The GRAPHICS option requests high-resolution graphics rather than character-based graphics.



If a new treatment has been applied to one group of people but no another, you may ask "Did the new treatment make a difference in the survival experience of the two groups of people?" A natural way is to compare the two survival curves between the two groups. We can simply add a STRATA statement after the TIME statement:

```
PROC LIFETEST data=myel plots=(s) graphics;
      TIME dur*status(0);
      STRATA treat;
RUN;
```



We observe that before 220 days the two curves are indistinguishable. The gap that develops after 220 days reflects the fact that no additional deaths occur in treatment group 1 after that time.

I strongly suggest you do the hand-calculation for the K-M estimates and sketch the survival plot based on your estimate by yourself and check if you did it right compared with the SAS plot.