BIOST P8110: Applied Regression II
5. PROC LIFETEST (Part I)

Qixuan Chen
Department of Biostatistics
Columbia University

This lecture's big ideas - how to use PROC LIFETEST in SAS to obtain:

1. Survival function estimate
2. Greenwood's standard error for survival function
3. 95% CI for survival function using log-log transformation
4. Estimates of quartiles and 95% CI
5. Estimate of mean and standard error
6. Plot of Kaplan-Meier survival curves

1. **Read in data set**

There are several ways to read data sets into SAS. Here, we introduce two approaches. The first approach is to type the data into SAS using DATA statement.

```
data data1;
input time status;
datalines;
10 1
13 1
13 0
14 1
17 1
19 0
23 1
25 0
;
run;
```

The second approach is to save the data set in an external file and then read the file into SAS. This approach is useful when the data set is large and used frequently.

```
data data1;
infile 'C:\cancer.csv' delimiter = ',' MISSOVER DSD;
input time status;
run;
```

In the "data1", the variable "time" is the survival time, and the variable "status" is censoring indicator with 1 for events and 0 for censored. Use MISSOVER if the last field or fields might be missing and you want SAS to assign missing values to the corresponding variable. DELIMITER specifies ', 'as the delimiter to be used for input. DSD (delimiter-sensitive data) specifies when data values are enclosed in quotation marks, delimiters within the values are treated as character data.

2. **SAS syntax**

LIFETEST procedure can be used to estimate survival functions, median survival time, and mean survival time.

```
proc lifetest data=data1 method=KM alpha=0.05 conftype=loglog outsurv=A
stderr;
time time*status(0);
run;
```

1) The PROC LIFETEST statement invokes the procedure.
    a. Optionally, this statement identifies an input "data=" and an output "outsurv=" data set.
    b. Optionally, this statement also specifies the computation details of the survival function estimation "method=", "alpha=", and "stderr".
    c. "method=" specifies the method used to compute the survival function estimates. They can be "KM" for Kaplan-Meier estimates or "LT" for life-table estimates.

d. "alpha=" specifies the level of significance for 100(1-α)% confidence interval.

e. "conftype=" specifies the transformation applied to S(t) to obtain the pointwise confidence intervals for the survival functions and the quantiles of the survival times.

f. "stderr" specifies the standard error of the survival function to be output to the "outsurv=" data set.

2) The TIME statement is used to specify the variables that define the survival time variable "time" and event or censoring indicator "status".

a. Because "0" is used to denote censored observation, a value of "0" is put inside the parenthesis after the censoring indicator "status".

b. If we use 1 to denote events and 2 to denote censored as in data2, a value of "2" will be used inside the parenthesis.

```
data data2;
input time status;
datalines;
10 1
13 1
13 2
14 1
17 1
19 2
23 1
25 2
;
Run;
proc lifetest data=data2 method=KM alpha=0.05 outsurv=A stderr;
time time*status(2);
run;
```

3. SAS output

The LIFETEST Procedure

Product-Limit Survival Estimates

time	Survival	Failure	Error	Number Failed	Number Left
0.0000	1.0000	0	0	0	8
10.0000	0.8750	0.1250	0.1169	1	7
13.0000	0.7500	0.2500	0.1531	2	6
13.0000*	.	.	.	2	5
14.0000	0.6000	0.4000	0.1817	3	4
17.0000	0.4500	0.5500	0.1882	4	3
19.0000*	.	.	.	4	2
23.0000	0.2250	0.7750	0.1849	5	1
25.0000*	.	.	.	5	0

NOTE: The marked survival times are censored observations.

Summary Statistics for Time Variable time

Quartile Estimates

Percent	Point Estimate	95% Confidence Interval Transform	[Lower	Upper)
75	23.0000	LOGLOG	14.0000	.
50	17.0000	LOGLOG	10.0000	.
25	13.5000	LOGLOG	10.0000	23.0000

Mean	Standard Error
17.8750	2.0566

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
8	5	3	37.50

```
proc print data=A; run;
```

The SAS System

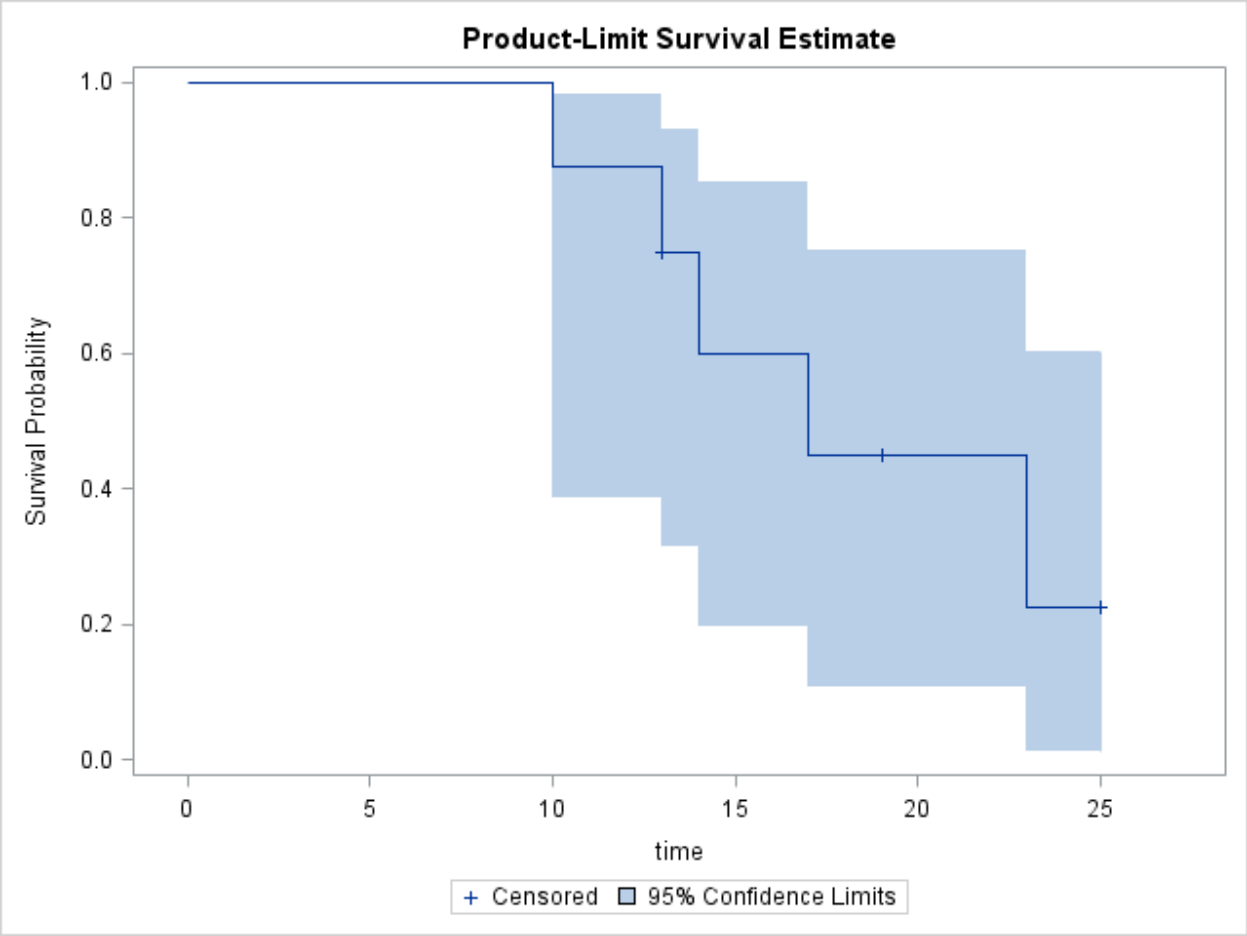
Obs	time	_CENSOR_	SURVIVAL	SDF_ STDERR	SDF_LCL	SDF_UCL
1	0	.	1.000	0.00000	1.00000	1.00000
2	10	0	0.875	0.11693	0.38700	0.98139
3	13	0	0.750	0.15309	0.31481	0.93090
4	13	1	0.750	.	.	.
5	14	0	0.600	0.18166	0.19550	0.85225
6	17	0	0.450	0.18825	0.10758	0.75127
7	19	1	0.450	.	.	.
8	23	0	0.225	0.18486	0.01240	0.60242
9	25	1	.	.	.	.

SURVIVAL: Kaplan-Meier estimate of survival function
SDF_STDERR: squared root of Greenwood's variance estimate of survival function
SDF_LCL and SDF_UCL: the lower and upper limits of the 95% CI for the survival functions based on the log-log transformation.

4. ODS Graphics for Kaplan-Meier survival plot

```
ods graphics on;
proc lifetest data=data1 method=KM alpha=0.05 maxtime=28 plots=survival(cl);
     time time*status(0);
run;
ods graphics off;
```

1) "maxtime=" specifies the maximum time value for x-axis.
2) "plots=" specifies the plots to display. "survival(cl)" specifies a plot of the estimated survival function versus time and the pointwise 95% CI.


Product-Limit Survival Estimate
Survival Probability
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25
time
+ Censored
95% Confidence Limits