

An Introduction to Event History Analysis

Oxford Spring School
June 18-20, 2007

Day One: Exploring Survival Data

Single-Record Data

The (completely made-up) example data:

```
. list id X T c
```

	id	X	T	c
1.	1	0	16	1
2.	2	0	17	1
3.	3	0	17	1
4.	4	0	21	0
5.	5	0	21	1
6.	6	0	27	0
7.	7	0	27	0
8.	8	0	29	1
9.	9	0	32	1
10.	10	0	33	1
11.	11	1	4	1
12.	12	1	7	1
13.	13	1	7	1
14.	14	1	12	1
15.	15	1	13	0
16.	16	1	15	1
17.	17	1	16	1
18.	18	1	17	1
19.	19	1	18	1
20.	20	1	21	0

```
. stset T, fail(c)
```

```
failure event: c != 0 & c < .  
obs. time interval: (0, T]  
exit on or before: failure
```

```
-----  
20 total obs.  
0 exclusions  
-----
```

```
20 obs. remaining, representing  
15 failures in single record/single failure data  
370 total analysis time at risk, at risk from t = 0  
earliest observed entry t = 0  
last observed exit t = 33
```

```
. stdes
```

```
failure _d: c  
analysis time _t: T
```

```
-----|----- per subject -----|  
Category          total      mean      min      median      max  
-----|-----  
no. of subjects          20  
no. of records          370      18.5      4      17      33  
  
(first) entry time          0      0      0      0  
(final) exit time      18.5      4      17      33  
  
subjects with gap          0  
time on gap if gap          0      .      .      .  
time at risk          370      18.5      4      17      33  
  
failures          15      .75      0      1      1  
-----|-----
```

Multiple-Record (Time-Varying) Data

The same data, only in time-varying format:

```
. list id duration censor
```

```

+-----+
| id   duration   censor |
+-----+
1. | 1         1         0 |
2. | 1         2         0 |
3. | 1         3         0 |
.   .         .         .
.   .         .         .
.   .         .         .
15. | 1        15         0 |
16. | 1        16         1 |
+-----+
17. | 2         1         0 |
18. | 2         2         0 |
.   .         .         .
.   .         .         .
.   .         .         .
33. | 2        17         1 |
34. | 3         1         0 |
35. | 3         2         0 |

```

etc.

```
. stset duration, fail(censor) id(id)
```

```

           id: id
failure event:  censor != 0 & censor < .
obs. time interval:  (duration[_n-1], duration]
exit on or before:  failure

```

```
370 total obs.
   0 exclusions
```

```
370 obs. remaining, representing
   20 subjects
   15 failures in single failure-per-subject data
370 total analysis time at risk, at risk from t =           0
           earliest observed entry t =           0
           last observed exit t =           33
```

. stdes

failure _d: censor
analysis time _t: duration
id: id

Category	total	----- per subject -----			
		mean	min	median	max
no. of subjects	20				
no. of records	370	18.5	4	17	33
(first) entry time		0	0	0	0
(final) exit time		18.5	4	17	33
subjects with gap	0				
time on gap if gap	0
time at risk	370	18.5	4	17	33
failures	15	.75	0	1	1

Counting-Process Data

```
. list id starttime endtime censor
```

```

+-----+
| id  startt~e  endtime  censor |
+-----+
1. | 1          0          1          0 |
2. | 1          1          2          0 |
3. | 1          2          3          0 |
.  | .          .          .          . |
.  | .          .          .          . |
.  | .          .          .          . |
15. | 1          14         15         0 |
16. | 1          15         16         1 |
+-----+
17. | 2          0          1          0 |
18. | 2          1          2          0 |
19. | 2          2          3          0 |
20. | 2          3          4          0 |
.  | .          .          .          . |
.  | .          .          .          . |
.  | .          .          .          . |
32. | 2          15         16         0 |
33. | 2          16         17         1 |
34. | 3          0          1          0 |
35. | 3          1          2          0 |
etc.

```

```
. stset endtime, fail(censor) enter(starttime) id(id)
```

```

          id:  id
failure event:  censor != 0 & censor < .
obs. time interval:  (endtime[_n-1], endtime]
enter on or after:  time starttime
exit on or before:  failure

```

```

370 total obs.
   0 exclusions

```

```

370 obs. remaining, representing
  20 subjects
  15 failures in single failure-per-subject data
370 total analysis time at risk, at risk from t =          0
          earliest observed entry t =          0
          last observed exit t =          33

```

. stdes

failure _d: censor
analysis time _t: endtime
enter on or after: time starttime
id: id

Category	total	----- per subject -----			
		mean	min	median	max
no. of subjects	20				
no. of records	370	18.5	4	17	33
(first) entry time		0	0	0	0
(final) exit time		18.5	4	17	33
subjects with gap	0				
time on gap if gap	0
time at risk	370	18.5	4	17	33
failures	15	.75	0	1	1

Basic Plots and Statistics

Figure 1: Kaplan-Meier Survival Function (with 95% Greenwood c.i.s)

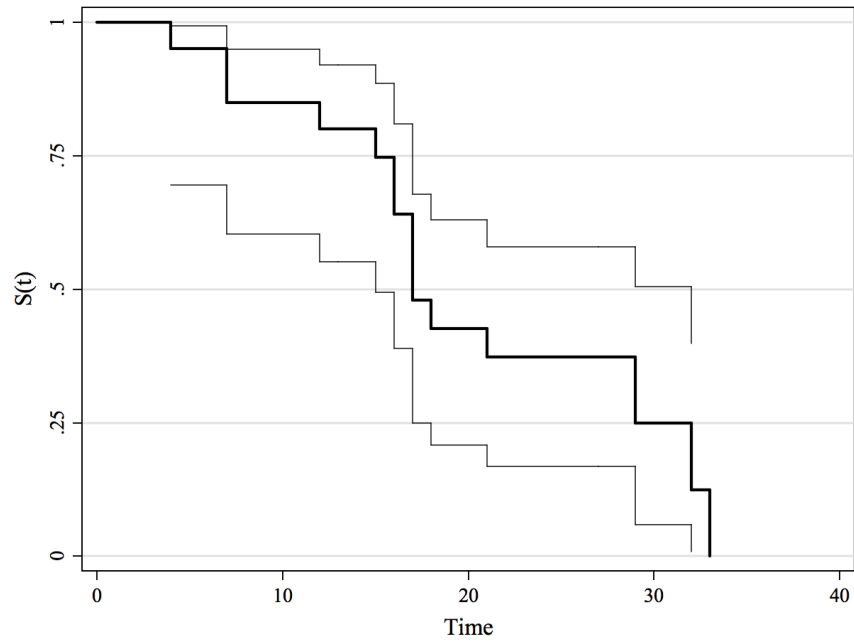
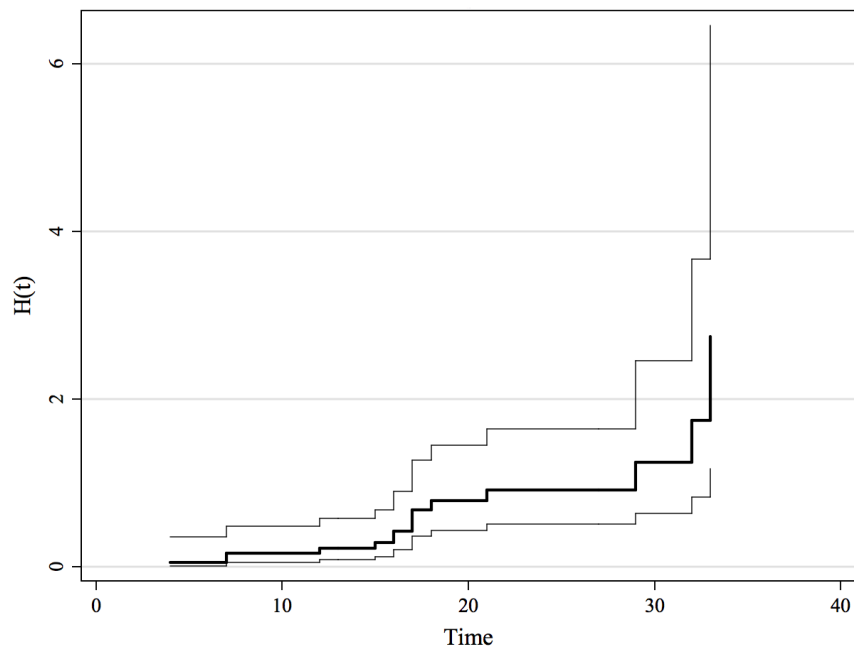


Figure 2: Nelson-Aalen Cumulative Hazard Function (with 95% c.i.s)



Compare two estimated survival curves for $X = 0$ and $X = 1$:

Figure 3: Kaplan-Meier Survival Function (with 95% Greenwood c.i.s), by X

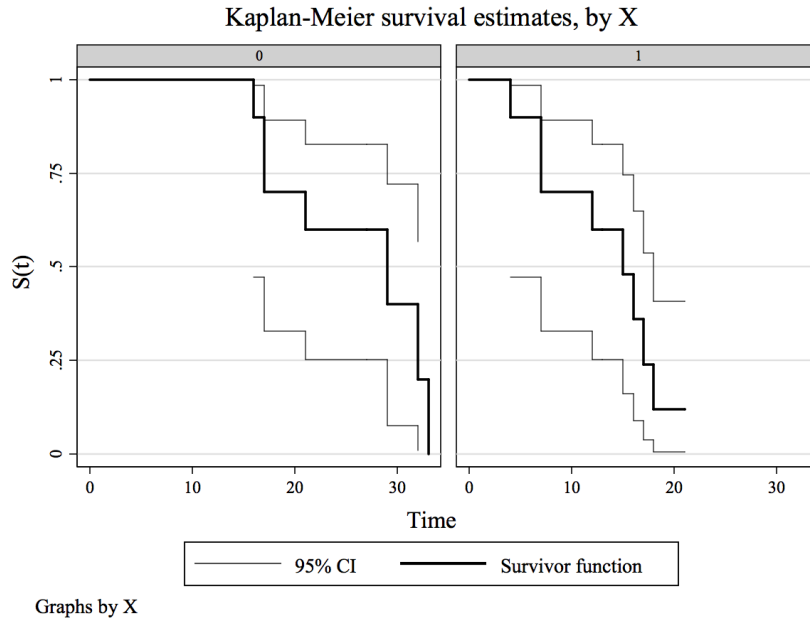
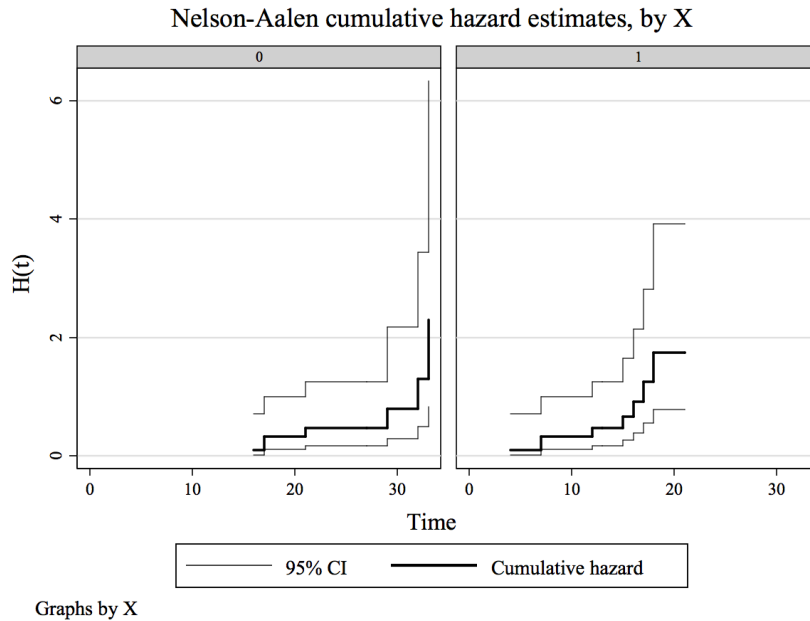


Figure 4: Nelson-Aalen Cumulative Hazard Function (with 95% c.i.s), by X



Log-Rank Test for the Equality of $\widehat{S}(t)$

```
. sts test X
```

```
          failure _d:  censor  
    analysis time _t:  endtime  
    enter on or after:  time starttime  
          id:  id
```

Log-rank test for equality of survivor functions

		Events	Events
X		observed	expected
X=0		7	11.14
X=1		8	3.86
Total		15	15.00

```
          chi2(1) =      7.43  
          Pr>chi2 =      0.0064
```

Survival Data in R

Reading in data:

```
Day2data<-read.dta(Day2data)  
Day2dataTV<-read.dta(Day2dataTV)
```

Creating survival objects:

```
NonTV<-Surv(Day2data$T, Day2data$c)  
TV<-Surv(Day2dataTV$starttime, Day2dataTV$endtime, Day2dataTV$censor)
```

Plotting a Kaplan-Meier Survival Function:

```
TVKM<-survfit(TV~1)  
plot(TVKM)
```

