An Applied Introduction

Christianna S. Williams
Abt Associates Inc, Durham, NC

Paper reprise presented at eSUG spring conference 2014 by

Niosha Gunasekara
Health Analytics Branch
Alberta Health
23rd April 2014
Overview

• Introduction
• Define Survival Time and Censoring
• Example:
  - Define Variables in SAS
  - Apply a Univariate Survival Method:
    *Kaplan-Meier Analysis*
  - Apply a Multivariate method:
    *Cox Regression* (a brief introduction)
• References
What is Survival Analysis?

Survival analysis is a family of statistical methods designed to analyze duration data (i.e., time until the occurrence of any well-defined event).

- Commonly used to evaluate time-to-event data in randomized clinical trials or cohort studies.

**Examples:**
- How long will cancer patients survive after a specific treatment?
- How long will it take graduate students to finish their degrees?
- What factors influence when children reach developmental milestones?

Survival time: a variable that measures the time from a particular starting Point to a particular endpoint of interest

- Start of treatment $\rightarrow$ Time of death
- Start of treatment $\rightarrow$ Development of functional ability
- Time of marriage $\rightarrow$ Birth of first child
Define Survival DATA

- Survival Data rarely normally distributed (skewed most of the time)

<table>
<thead>
<tr>
<th>Complete data</th>
<th>Left censored</th>
</tr>
</thead>
<tbody>
<tr>
<td>value of each sample unit is observed or known.</td>
<td>know only the maximum value</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Interval Censored</th>
<th>Right censored</th>
</tr>
</thead>
<tbody>
<tr>
<td>lie between a certain minimum and maximum</td>
<td>know only the minimum value</td>
</tr>
</tbody>
</table>

Sources:
http://www.vosesoftware.com/ModelRiskHelp/index.htm#Analysing_and_using_data/Fitting_distributions/Censored_data.htm
Define Censoring

Censor if the exact time at which the event occurs is not observed

Censor if a participant drop out or die

Figure 1. Hypothetical survival data for six patients. See text for further description.
Example: Lachs & Williams et al. (1997)

- **Research Question:** Were mistreated older adults more likely to be admitted to a Nursing Home (or admitted sooner) compared to those who were not mistreated?

- **Notes:** Adult mistreatment in this study was treated as time dependent covariates and defined based on exposure to either of the following conditions:
  i) External cause of elder abuse
  ii) Self neglect
Example

• Used two survival methods:

1) Kaplan Meier analysis to compute the probability of NH admission as a function of time and compare differences in survival probabilities for gender and marital status

2) Cox regression analysis to examine the effect of many variables including time-dependent covariates on hazard function

Cohort: 2800 patients >65 yrs lived in New Haven, CN – enrolled in a large study of aging in Feb 1982 – followed up until Dec 1995
Example: Variables

- **Study enrollment (basedate)**: 1982-02-28 to 1982-12-31
- **Study end date (endfwpdate)**: 1995-12-31
- **NH Admit indicator (nhadmit)**: 0 or 1 indicator of whether the person had a NH admission during the follow up
- **NH Admit date (nhpdate)**: Date first admitted to a NH
- **Death indicator (died)**: 0 or 1 indicator of death during follow up
- **Date of Death (deathdate)**: Date that the person died
- **Date Censored (censdate)**: Date that an observation is censored
- **Survival Time (eventdys)** = (censdate) – (basedate)
1 - Define Variables

**Study enrollment**

\( \text{basedate} \): 1982-12-31

Admitted to a NH within the follow up period 😊

Censor=0

censdate=nhdate

Patient died during the follow up 😔

Censor=1

censdate=deathdate

Patient survived beyond follow up 😞

Censor=2

censdate=endfwpdt

**Study end date**

\( \text{endfwpdate} \): 1995-12-31

😊Failed (censor=0)

😔 Censored (censor=1 or 2)
Create Variables using SAS

**Study enrollment**
(basedate): 1982-12-31

**Study end date**
(endfwpdate): 1995-12-31

Event Occurs 😊
Admitted to a NH within the follow up period
Censor=0
censdate=nhdate

Patient died during the follow up 😞
Censor=1
censdate=deathdate

Patient survived beyond follow up 😊
Censor=2
censdate=endfwpdt

endfwpdate = MDY(12,31,1995);

IF (nhadmit = 1) AND (basedate LE nhdate LE endfwpdt) THEN DO;
censor = 0;
censdate = nhdate ;
END;

ELSE IF (died = 1) AND (basedate LE deathdate LE endfwpdt) THEN DO;
censor = 1;
censdate = deathdate ;
END;

ELSE IF (died NE 1) OR (deathdate GT endfwpdt) THEN DO;
censor = 2;
censdate = endfwpdt ;
END;

** time on study -- baseline to nh admit/death/end of study ;
eventdys = censdate - basedate ;
Kaplan-Meier survival curves

- Also known as product-limit formula and accounts for censoring

- Totally nonparametric. No assumptions about the underlying true distribution of failure times.

- Does not account for confounding or effect modification by other covariates

- Survival estimates can be unreliable toward the end of a study, when there are small numbers of subjects at risk of having an event

An example to show censored observations in a KM plot

Source: http://cancerguide.org/scurve_km.html
3 – Plot Kaplan Meier Survival Curves

**PROC LIFETEST**
Plots the survival distribution function, using the Kaplan-Meier method.

**Syntax:**

```sas
PROC LIFETEST DATA = em_nh1 METHOD=KM PLOTS=S CS=none;
TIME eventdys*censor(1,2) ;
TITLE1 FONT="Arial 10pt" HEIGHT=1 BOLD 'Kaplan-Meier Curve --overall' ;
RUN;
```

- **Specify data**
- **Tells SAS which method to use for life test procedure**
- **Censor Symbol**
- **Tells SAS which values are censored**
- **Tells SAS to create the Kaplan-Meier estimate survival plots**
- **tells SAS the Survival time (event time) variable**
The study didn’t last until the median survival time (i.e. fewer than half had been placed in a nursing home by the end of the study).
### Kaplan Meier Analysis

#### Time = 0 Days

- Everyone is surviving.

#### Time = 4978 Days

- Probability of survival = 0.5256

---

**Output 1. Subset of Product Limit (aka Kaplan Meier) Estimates**

<table>
<thead>
<tr>
<th>EVENTDYS</th>
<th>Survival</th>
<th>Standard Error</th>
<th>Failure</th>
<th>Standard Error</th>
<th>Number</th>
<th>Failed</th>
<th>Left</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>1.0000</td>
<td>0.361</td>
<td>0</td>
<td>0.000361</td>
<td>0</td>
<td>0</td>
<td>2769</td>
</tr>
<tr>
<td>2.00*</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>0</td>
<td>0</td>
<td>2768</td>
</tr>
<tr>
<td>6.00</td>
<td>0.9996</td>
<td>0.000361</td>
<td>0.000361</td>
<td></td>
<td>1</td>
<td>1</td>
<td>2767</td>
</tr>
<tr>
<td>7.00</td>
<td>0.9989</td>
<td>0.00108</td>
<td>0.000625</td>
<td></td>
<td>2</td>
<td>2</td>
<td>2766</td>
</tr>
<tr>
<td>7.00</td>
<td>0.9989</td>
<td>0.00108</td>
<td>0.000625</td>
<td></td>
<td>3</td>
<td>3</td>
<td>2765</td>
</tr>
<tr>
<td>18.00*</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>3</td>
<td>3</td>
<td>2764</td>
</tr>
<tr>
<td>19.00*</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>3</td>
<td>3</td>
<td>2763</td>
</tr>
<tr>
<td>22.00*</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>3</td>
<td>3</td>
<td>2762</td>
</tr>
<tr>
<td>25.00</td>
<td>0.9986</td>
<td>0.00145</td>
<td>0.000722</td>
<td></td>
<td>4</td>
<td>4</td>
<td>2761</td>
</tr>
<tr>
<td>26.00*</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>4</td>
<td>4</td>
<td>2760</td>
</tr>
<tr>
<td>28.00*</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>4</td>
<td>4</td>
<td>2759</td>
</tr>
<tr>
<td>32.00</td>
<td>0.9982</td>
<td>0.00181</td>
<td>0.000808</td>
<td></td>
<td>5</td>
<td>5</td>
<td>2758</td>
</tr>
<tr>
<td>33.00*</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>5</td>
<td>5</td>
<td>2757</td>
</tr>
<tr>
<td>33.00*</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>5</td>
<td>5</td>
<td>2756</td>
</tr>
</tbody>
</table>

---

Note: The marked survival times are censored observations.
3 – SAS Output: KM Analysis cont....

Out of all, 25% of participants had had an event by 2,512 days.

The study didn’t last until the median survival time (i.e. fewer than half had been placed in a nursing home by the end of the study).

A total of 935 people had an event and 1834 were censored (either died during follow-up without having entered a nursing home or were alive and not in a nursing home when the study ended).
4 – Add Strata to KM Analysis

**Syntax:**

```plaintext
PROC LIFETEST DATA = em_nh1 METHOD=KM PLOTS=S CS=NONE;
TIME eventdys*censor(1,2) ;
STRATA maried82 ;
SYMBOL1 V=none COLOR=blue LINE=1;
SYMBOL2 V=none COLOR=red LINE=2;
RUN;
```

Allow comparing survival probabilities for two groups (Married and Unmarried)
Output 6. Kaplan-Meier survival curves according to marital status

Kaplan-Meier Curve -- by Marital Status

Days from baseline to NH admission/Death/end-of-follow-up
4 - Test of Equality over Strata

• **Hypotheses:**
  
  $H_0 =>$ the risk of the groups are equal  
  $H_a =>$ the risk of the groups are not equal

• **The Wilcoxon test** - more powerful in detecting differences **earlier** in follow up

• **The Log-Rank test** - more powerful in detecting differences **later** in follow up
5 – Cox Regression

- Multivariate survival technique for time-to-event data based on hazard function
- Produces multivariate-adjusted hazard ratios
- Can be used to account for time-dependent covariates if proportional hazard requirement is violated

PROC PHREG DATA = em_nh1 ;
CLASS gender ;
MODEL eventdys*censor(1,2) = gender age82 maried82 bmi82 cesd82 /RL;
RUN;
5 – Coding and Testing for Time Dependent covariates using Cox Regression

Figure 2. Schematic depiction of survival data for six patients, with a time dependent covariate. See text for further description.
5 – Coding and Testing for Time Dependent covariates

PROC PHREG DATA = em_nh1 ;
CLASS GENDER ;
MODEL eventdys*censor(1,2) = vems vsn gender age82 maried82 bmi82 cesd82 /RL TIES=EFRON;

IF (0 LE vemsdays LE eventdys) THEN DO;
vems = 1;
vsn = 0;
END;
ELSE vems = 0;

IF vems NE 1 THEN DO;
IF (0 LE vslfdays LE eventdys) THEN vsn = 1;
ELSE vsn = 0;
END;
RUN;

Vems = Verified case of elder mistreatment
Vsn = Verified case of self-neglect
**Output 11. PHREG output with a time-dependent covariate**

### Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
<th>Hazard Ratio</th>
<th>95% Hazard Ratio Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>vems</td>
<td>1</td>
<td>1.17229</td>
<td>0.23621</td>
<td>24.6301</td>
<td>&lt;.0001</td>
<td>3.229</td>
<td>2.033</td>
</tr>
<tr>
<td>vsn</td>
<td>1</td>
<td>1.82335</td>
<td>0.13393</td>
<td>185.3456</td>
<td>&lt;.0001</td>
<td>6.193</td>
<td>4.763</td>
</tr>
<tr>
<td>gender F</td>
<td>1</td>
<td>-0.15470</td>
<td>0.08021</td>
<td>3.7200</td>
<td>0.0538</td>
<td>0.857</td>
<td>0.732</td>
</tr>
<tr>
<td>AGE82</td>
<td>1</td>
<td>0.09040</td>
<td>0.00511</td>
<td>313.4993</td>
<td>&lt;.0001</td>
<td>1.095</td>
<td>1.084</td>
</tr>
<tr>
<td>MARRIED82</td>
<td>1</td>
<td>-0.32184</td>
<td>0.08775</td>
<td>13.4519</td>
<td>0.0002</td>
<td>0.725</td>
<td>0.610</td>
</tr>
<tr>
<td>BMI82</td>
<td>1</td>
<td>-0.02272</td>
<td>0.00850</td>
<td>7.1513</td>
<td>0.0075</td>
<td>0.978</td>
<td>0.961</td>
</tr>
<tr>
<td>CESD82</td>
<td>1</td>
<td>0.02229</td>
<td>0.00379</td>
<td>34.6301</td>
<td>&lt;.0001</td>
<td>1.023</td>
<td>1.015</td>
</tr>
</tbody>
</table>
References

...and here's a chart that shows what you might see if you looked at a mountain range through a tennis racket.