Survival Analysis using SAS

Rajeev Kumar
Fisheries Center, UBC, Vancouver
Email: r.kumar AT live.com

Divya Varkey
Fisheries Center, UBC, Vancouver
Email: d.varkey AT live.com

Vancouver SAS Users Group meeting
May 30th, 2012
Outline

- What is Survival Analysis
  - Data description

- Univariate Analysis
  - Kaplan-Meier method
    - Survival curve and log-rank test

- Multivariate Analysis
  - Cox Proportional Hazard (PH) model
    - Model selection
    - PH assumption
    - Modelling: time-dependent covariates
Introduction: dealing with time-event data

- Censored values makes the analysis complex
  - right censor is most common
Data description: myeloma, 65 patients

```sas
proc print data=myeloma (where=(ranuni(1)<=.15)); run;
```

<table>
<thead>
<tr>
<th>Obs</th>
<th>Time</th>
<th>Alive_dead</th>
<th>Age</th>
<th>SEX</th>
<th>LogBUN</th>
<th>HGB</th>
<th>Platelet</th>
<th>Fracture</th>
<th>LogPBM</th>
<th>Infection</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>6</td>
<td>1</td>
<td>77</td>
<td>1</td>
<td>1.3617</td>
<td>9.0</td>
<td>1</td>
<td>0</td>
<td>1.4624</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>6</td>
<td>1</td>
<td>70</td>
<td>2</td>
<td>2.1139</td>
<td>10.2</td>
<td>0</td>
<td>1</td>
<td>1.3617</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>7</td>
<td>1</td>
<td>61</td>
<td>2</td>
<td>1.0414</td>
<td>5.1</td>
<td>0</td>
<td>1</td>
<td>2.0000</td>
<td>0</td>
</tr>
<tr>
<td>47</td>
<td>89</td>
<td>1</td>
<td>63</td>
<td>1</td>
<td>1.3222</td>
<td>14.0</td>
<td>1</td>
<td>1</td>
<td>1.6232</td>
<td>1</td>
</tr>
<tr>
<td>48</td>
<td>92</td>
<td>1</td>
<td>58</td>
<td>2</td>
<td>1.4314</td>
<td>11.0</td>
<td>1</td>
<td>1</td>
<td>1.4150</td>
<td>0</td>
</tr>
<tr>
<td>54</td>
<td>12</td>
<td>0</td>
<td>46</td>
<td>2</td>
<td>1.1461</td>
<td>11.6</td>
<td>1</td>
<td>0</td>
<td>1.1461</td>
<td>0</td>
</tr>
<tr>
<td>60</td>
<td>19</td>
<td>0</td>
<td>69</td>
<td>2</td>
<td>1.3222</td>
<td>10.8</td>
<td>1</td>
<td>1</td>
<td>1.5185</td>
<td>0</td>
</tr>
<tr>
<td>63</td>
<td>53</td>
<td>0</td>
<td>66</td>
<td>1</td>
<td>1.1139</td>
<td>12.0</td>
<td>1</td>
<td>1</td>
<td>2.0000</td>
<td>0</td>
</tr>
</tbody>
</table>

Source: Krall et al. (1975); published in *Biometrics* (Vol. 31, No 1, pp. 49-57)
Survival probability and hazard rate

- Kaplan-Meier (K-M) survival probability
  \[ S_t = (S_{t-1}) \times (1 - \frac{d_t}{n_t}) \]

- Hazard rate: risk of failure at time \( t \)
  \[ \lambda_t = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \]
SAS: commonly used approach

- Proc lifetest:

- Proc phreg:

- Proc lifereg:
  - for left, right, uncensored
  - it has options for define distribution for survival time (such as exponential, gamma, weibull, normal etc.)
Univariate analysis: proc lifetest

```plaintext
proc format;
  value sex 1="M" 2="F";
  value platelet 0="Abnormal" 1="Normal";
  value fracture 0="No" 1="Yes";
  value infection 0="No" 1="Yes";
  value logbun low - 1.5 = "<1.5" 1.5 - 2.0 = "1.5-2.0" 2 - high = ">2";
  value hgb low - 10 = "<10" 10 - high = ">10";
  value logpbm low - 1.5 = "<1.5" 1.5 - high = ">1.5";
  value age low - 50 = "<50" 50 - 60 = "50-60" 60 - high = ">60";
run;

proc lifetest data=myeloma plots=s (test cl atrisk);
  time time*alive_dead(0);
  strata logbun;
  format logbun logbun.;
run;
```

![Graph showing product-limit survival estimates with number of subjects at risk and 95% confidence limits.](image)
**Contd..**

```plaintext
proc lifetest data=myeloma plots=(s lls);
  time time*alive_dead(0);
  strata hgb;
  format hgb hgb.;
run;
```
Univariate analysis: K-M curve
K-M curve contd..

Product-Limit Survival Estimates

- Censored
- Logrank p=0.2879

Survival Probability

Time

Infection
- No
- Yes

Product-Limit Survival Estimates

- Censored
- Logrank p=0.0696

Survival Probability

Time

Platelet
- Abnormal
- Normal

Product-Limit Survival Estimates

- Censored
- Logrank p=0.0076

Survival Probability

Time

HGB
- <10
- >10

Product-Limit Survival Estimates

- Censored
- Logrank p < 0.001

Survival Probability

Time

LogBUN
- 1.5-2.0
- <1.5
- >2

30-May-2012    VanSUG
Multivariate analysis
Cox proportional hazards regression model

- \( h(t|X) = h_0(t) \cdot \exp(X_1 \beta_1 + X_2 \beta_2 + \cdots + X_n \beta_n) \)

Baseline hazard | Exponential part

- Base line hazard: function of time
- Exponential part: time independent

Hazard Ratio (HR) = \( \frac{\text{est. } h(t|X_1=1)}{\text{est. } h(t|X_1=0)} \)

- PH assumption: hazard between group is proportional over time
- Proportionality constant is time independent
Cox proportional hazard model

```
proc phreg data=myeloma;

format sex sex. platelet platelet.
      fracture fracture. infection infection. ;

class sex (ref="M") platelet (ref="Normal") fracture (ref="No")
     infection (ref="No") ;

model time*alive_dead (0)= sex platelet fracture infection age logpbm
     logbun  hgb / ties=efron ;

run;
```
Model Selection

- Backward elimination
- Forward selection
- Stepwise

- Finally we use BIC or AIC values for determining best model
Model selection contd..

```sas
proc phreg data=myeloma;
format sex sex. platelet platelet. fracture fracture. infection infection. ;
class sex (ref="M") platelet (ref="Normal") fracture (ref="No") infection (ref="No") ;
model time*alive_dead (0)= sex platelet fracture infection age logpbm logbun hgb / ties=efron
  selection=s slentry=0.2 slstay=0.05 ;
run;
```

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Without Cov</th>
<th>LogBUN</th>
<th>LogBUN+HGB</th>
<th>LogBUN+HGB+Infection</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2 LOG L</td>
<td>308.389</td>
<td>300.374</td>
<td>296.121</td>
<td>294.263</td>
</tr>
<tr>
<td>AIC</td>
<td>308.389</td>
<td>302.374</td>
<td>300.121</td>
<td>300.263</td>
</tr>
<tr>
<td>SBC</td>
<td>308.389</td>
<td>304.245</td>
<td>303.864</td>
<td>305.876</td>
</tr>
</tbody>
</table>
Ph Assumption

Kolmogorov-Type Supremum Test for Proportional Hazards

```sas
proc phreg data=myeloma;
model time*alive_dead (0)= logbun hgb;
assess ph/resample;
run;
```

Checking Proportional Hazards Assumption for LogBUN
Observed Path and First 20 Simulated Paths

Checking Proportional Hazards Assumption for HGB
Observed Path and First 20 Simulated Paths

Pr > MaxAbsVal: 0.0560
(1000 Simulations)

Pr > MaxAbsVal: 0.5580
(1000 Simulations)
**PH assumption contd..**

```r
proc phreg data=myeloma;
model time*alive_dead (0)= logbun  hgb logbunT;
logbunT=logbun*log(time);
*phtest: test logbunT;
run;
```

<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>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>LogBUN</td>
<td>1</td>
<td>4.67223</td>
<td>1.32403</td>
<td>12.4525</td>
<td>0.0004</td>
<td>106.936</td>
<td>LogBUN</td>
</tr>
<tr>
<td>HGB</td>
<td>1</td>
<td>-0.13256</td>
<td>0.05914</td>
<td>5.0246</td>
<td>0.0250</td>
<td>0.876</td>
<td>HGB</td>
</tr>
<tr>
<td>logbunT</td>
<td>1</td>
<td>-1.52671</td>
<td>0.59302</td>
<td>6.6279</td>
<td>0.0100</td>
<td>0.217</td>
<td></td>
</tr>
</tbody>
</table>
**Time-dependent covariate**

```sas
proc phreg data=myeloma;
  logtime=log(time);
  model time*alive_dead (0)= logbun  hgb logbun*logtime /risklimits;
  hazardratio 'logBUN' logbun / at(logtime=0,
               1.7917, 2.4849, 3.1780, 3.5835, 3.8712) cl=both;
run;
```

**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>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>LogBUN</td>
<td>1</td>
<td>4.67223</td>
<td>1.32403</td>
<td>12.4525</td>
<td>0.0004</td>
<td>.</td>
<td>.</td>
<td>LogBUN</td>
</tr>
<tr>
<td>HGB</td>
<td>1</td>
<td>-0.13256</td>
<td>0.05914</td>
<td>5.0246</td>
<td>0.0250</td>
<td>0.876</td>
<td>0.780 0.983</td>
<td>HGB</td>
</tr>
<tr>
<td>logtime*LogBUN</td>
<td>1</td>
<td>-1.52671</td>
<td>0.59302</td>
<td>6.6279</td>
<td>0.0100</td>
<td>.</td>
<td>.</td>
<td>logtime * LogBUN</td>
</tr>
</tbody>
</table>

**logBUN: Hazard Ratios for LogBUN**

<table>
<thead>
<tr>
<th>Description</th>
<th>Point Estimate</th>
<th>95% Wald Confidence Limits</th>
<th>95% Profile Likelihood Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>LogBUN Unit=1 At logtime=0</td>
<td>106.936</td>
<td>7.982 1432.644</td>
<td>8.594 1652.555</td>
</tr>
<tr>
<td>LogBUN Unit=1 At logtime=1.7917</td>
<td>6.937</td>
<td>2.008 23.962</td>
<td>2.004 24.264</td>
</tr>
<tr>
<td>LogBUN Unit=1 At logtime=2.4849</td>
<td>2.407</td>
<td>0.617 9.390</td>
<td>0.600 9.179</td>
</tr>
<tr>
<td>LogBUN Unit=1 At logtime=3.178</td>
<td>0.836</td>
<td>0.130 5.378</td>
<td>0.124 5.177</td>
</tr>
<tr>
<td>LogBUN Unit=1 At logtime=3.5835</td>
<td>0.450</td>
<td>0.048 4.222</td>
<td>0.045 4.021</td>
</tr>
<tr>
<td>LogBUN Unit=1 At logtime=3.8712</td>
<td>0.290</td>
<td>0.023 3.625</td>
<td>0.022 3.424</td>
</tr>
</tbody>
</table>
The analysis shown in this presentation is only for tutorial purpose.