Lab Ten: Survival Analysis II: Cox Regression

Lab Objectives
After today’s lab you should be able to:

1. Practice dealing with SAS date/time variables.
2. Fit models using PROC PHREG. Understand PROC PHREG output.
3. Understand output from the “baseline” statement.
4. Output estimated survivor functions and plot cumulative hazards.
5. Understand the role of the strata statement in PROC PHREG.
6. Evaluate PH assumption graphically.
7. Output and plot predicted survivor functions at user-specified levels of the covariates.
Follow along with the computer in front…

1. Double-click on to open the SAS editor file “data creation code” which should be saved in your stats210 folder from last week; run the libname statement:

```sas
libname stats210 'C:\Documents and Settings\mitl-pc.LANE-LIB\My Documents\Stats210';
```

2. Using the Explorer Browser on the left hand side of your screen, double check that a stats210 library has been properly created, and that it contains the SAS dataset “runners”

3. Plot the Kaplan-Meier survival curve for the by level of daily calcium intake (<800 mg/day, 800-1499 mg/day, and 1500+ mg day) using the strata statement as below:

```sas
proc lifetest data=stats210.runners plots=(s) graphics censoredsymbol=none maxtime=24;
  time time*sfl(0);
  strata calc1(800,1500);
  title 'Kaplan-Meier plot of fracture-free survivorship by calcium level';
  symbol1 v=none c=black w=2 i=join line=1;
  symbol2 v=none c=black w=2 i=join line=2;
  symbol3 v=none c=black w=2 i=join line=3;
 run;
```

This asks SAS to divide into groups as follows: \([-\infty,800) \[800,1500) \[1500,\infty)\]. This is an extremely useful feature of PROC LIFETEST, because you don’t have to break the variables into categories yourself; SAS does it for you.

With small numbers in each group, this is a pretty striking difference in fracture rates by calcium level, though it is not statistically significant yet (p=.13). A significant log-rank test here would mean that at least one pair of groups is significantly different in fracture rate.

<table>
<thead>
<tr>
<th>Test of Equality over Strata</th>
<th>Chi-Square</th>
<th>DF</th>
<th>Pr &gt; Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log-Rank</td>
<td>3.9352</td>
<td>2</td>
<td>0.1398</td>
</tr>
<tr>
<td>Wilcoxon</td>
<td>4.0947</td>
<td>2</td>
<td>0.1291</td>
</tr>
<tr>
<td>-2Log(LR)</td>
<td>4.5365</td>
<td>2</td>
<td>0.1035</td>
</tr>
</tbody>
</table>
4. Last time, we saw that there was no difference in fracture rates between those randomized to control and those randomized to treatment (so called “intention-to-treat analysis”). We might also want to know if there is a difference in fracture rates between those who actually took oral contraceptives and those who did not. We recorded the start and stop dates of taking treatment in the date variables “startoc” and “endoc.” Use these variables to create a new variable that represents the amount of time (in months) that each woman actually took the treatment (regardless of her assignment to treatment or control). Recall that date variables in SAS are recorded as the number of days since Jan 1., 1960.

```sas
data stats210.runners;
set stats210.runners;
OCTIME=(endoc-startoc)/365.25*12;
run;
```

5. Use: SolutionsÆAnalysisÆInteractive Data AnalysisÆStats210ÆRunnersÆOCTIMEÆAnalyzeÆDistributionÆTablesÆFrequency counts, to examine the distribution of time-on-treatment for the study participants. Notice it’s basically bimodal—women either took OCs for 0 or very few months or for more than 1 year.

6. Plot a Kaplan-Meier Curve that compares women who took OC’s for at least 6 months against women who never took them:

```sas
proc lifetest data=stats210.runners plots=(s) graphics
censoredsymbol=none;
time time*sf1(0);
title 'Kaplan-Meier plot of Fracture-free survivorship';
strata octime (6);
run;
```

7. Use Cox Regression to examine the relationship between treatment use (actual) and treatment randomization (intention-to-treat) and generate hazard ratios with confidence limits:

```sas
proc phreg data=stats210.runners;
model time*sf1(0)=octime / risklimits;
title 'Cox model for runners data—actual OC time';
run;
```

```sas
proc phreg data=stats210.runners;
model time*sf1(0)=treatr / risklimits;
title 'Cox model for runners data—treatment randomization';
run;
```

Asks for 95% confidence limits for the hazard ratios.
Examine Output:
Cox model for runners data-actual OC time

The PHREG Procedure

Model Information

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Dependent Variable</th>
<th>Censoring Variable</th>
<th>Censoring Value(s)</th>
<th>Ties Handling</th>
</tr>
</thead>
<tbody>
<tr>
<td>STATS210.RUNNERS</td>
<td>time</td>
<td>sf1</td>
<td>0</td>
<td>BRESLOW</td>
</tr>
</tbody>
</table>

Model Fit Statistics

<table>
<thead>
<tr>
<th>Criterion Type</th>
<th>Without Covariates</th>
<th>With Covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2 LOG L</td>
<td>67.500</td>
<td>67.444</td>
</tr>
<tr>
<td>AIC</td>
<td>67.500</td>
<td>69.444</td>
</tr>
<tr>
<td>SBC</td>
<td>67.500</td>
<td>69.641</td>
</tr>
</tbody>
</table>

Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Testing Global Null Hypothesis: BETA=0

<table>
<thead>
<tr>
<th>Test</th>
<th>Chi-Square</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likelihood Ratio</td>
<td>0.0555</td>
<td>1</td>
<td>0.8138</td>
</tr>
<tr>
<td>Score</td>
<td>0.0552</td>
<td>1</td>
<td>0.8142</td>
</tr>
<tr>
<td>Wald</td>
<td>0.0551</td>
<td>1</td>
<td>0.8144</td>
</tr>
</tbody>
</table>

Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter DF</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>OCTIME</td>
<td>1</td>
<td>-0.00651</td>
<td>0.02772</td>
<td>0.0551</td>
<td>0.994</td>
<td>0.941 1.049</td>
</tr>
</tbody>
</table>

0.6% decrease in hazard rate (=instantaneous risk of fracture) for every 1-month increase in OC use. Very close to the null value.
For treatment randomization:

Testing Global Null Hypothesis: BETA=0

<table>
<thead>
<tr>
<th>Test</th>
<th>Chi-Square</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likelihood Ratio</td>
<td>0.0193</td>
<td>1</td>
<td>0.8895</td>
</tr>
<tr>
<td>Score</td>
<td>0.0194</td>
<td>1</td>
<td>0.8893</td>
</tr>
<tr>
<td>Wald</td>
<td>0.0194</td>
<td>1</td>
<td>0.8893</td>
</tr>
</tbody>
</table>

Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Standard</th>
<th>Hazard</th>
<th>95% Hazard Ratio</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>treate</td>
<td>0.0936</td>
<td>1.098</td>
<td>0.295</td>
<td>TREATR</td>
</tr>
</tbody>
</table>

9% increase in hazard for being randomized to treatment. Also close to the null value, and similar to Odds Ratio results from logistic regression and contingency tables.

8. There doesn’t seem to be much there. But we might think that the relationship between actual time on OC’s and fracture protection is not linear. Could try OC use as a binary predictor (as in step 5). Note the convenient ability to create new (temporary) variables within PROC PHREG:

```sas
proc phreg data=stats210.runners;
  model time*sf1(0)=oc / risklimits;
  if octime>=6 then oc=6 ;
  if octime<6 then oc=0 ;
  title 'Cox model for runners data-OC use >=6 month';
run;
```

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter</th>
<th>Standard</th>
<th>Hazard</th>
<th>95% Hazard Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>oc</td>
<td>-0.04785</td>
<td>0.11183</td>
<td>0.953</td>
<td>0.766</td>
</tr>
</tbody>
</table>

Essentially no difference.
9. Another convenient feature of PROC PHREG is that you can allow that the baseline hazard might be different across different groups, and you can stratify on different groups. For example, in the runners study there is reason to believe that the athletes recruited from different sites might have very different characteristics (some sites recruited mostly from collegiate runners and others recruited only from post-collegiate runners) and might have very different baseline fracture rates. Therefore, we can stratify on fracture rates as follows:

```sas
proc phreg data=stats210.runners;
  model time*sfl(0)=oc / risklimits;
  if octime>=6 then oc=6 ;
  if octime<6 then oc=0 ;
  strata sitenum;
  title 'Cox model for runners data-OC use >=6 month';
run;
```

10. Evaluate the effects of several predictors, each adjusted for each other:

```sas
title ' ';
proc phreg data=stats210.runners;
  model time*sfl(0)=calc1 stressf bmc1 edever menarch / risklimits;
  baseline out=outdata survival=S;
run;
proc print data=outdata;
run;
```

<table>
<thead>
<tr>
<th>Test</th>
<th>Chi-Square</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likelihood Ratio</td>
<td>17.5738</td>
<td>5</td>
<td>0.0035</td>
</tr>
<tr>
<td>Score</td>
<td>18.3211</td>
<td>5</td>
<td>0.0026</td>
</tr>
<tr>
<td>Wald</td>
<td>11.3943</td>
<td>5</td>
<td>0.0441</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
<th>Hazard Ratio</th>
<th>95% Hazard Ratio Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>calc1</td>
<td>1</td>
<td>-0.00172</td>
<td>0.0007759</td>
<td>4.8936</td>
<td>0.0270</td>
<td>0.998</td>
<td>0.997 - 1.000</td>
</tr>
<tr>
<td>stressf</td>
<td>1</td>
<td>1.90860</td>
<td>0.99539</td>
<td>3.6766</td>
<td>0.0552</td>
<td>6.744</td>
<td>0.959 - 47.443</td>
</tr>
<tr>
<td>bmc1</td>
<td>1</td>
<td>-0.00118</td>
<td>0.00109</td>
<td>1.1765</td>
<td>0.2781</td>
<td>0.999</td>
<td>0.997 - 1.001</td>
</tr>
<tr>
<td>edever</td>
<td>1</td>
<td>5.64477</td>
<td>2.31313</td>
<td>5.9552</td>
<td>0.0147</td>
<td>282.809</td>
<td>3.038 - 26328.80</td>
</tr>
<tr>
<td>menarch</td>
<td>1</td>
<td>-0.64796</td>
<td>0.34077</td>
<td>3.6155</td>
<td>0.0572</td>
<td>0.523</td>
<td>0.268 - 1.020</td>
</tr>
</tbody>
</table>

Controlling for other variables in the model, calcium significantly decreases fracture risk by more than 80% for every 1000mg/day: 

\[ \beta_{calc} = -0.00172 \]

\[ \beta_{calc}(1000mg) = -0.00172(1000) = -1.72 \]

\[ HR_{1000mg} = e^{-1.72} = 0.179 \]
11. SAS can also calculate a baseline survivor function (recall that the baseline survivor function is NOT estimated by Cox regression). SAS uses a non-parametric method to estimate the baseline survivor function. To get the estimated survivor function not accounting for any covariates (similar to Kaplan-Meier):

```sas
title ' '; 
model time*sfl(0)=/ risklimits;
baseline out=outdata survival=S;
run;
proc print data=outdata;
run;
axis1 order=(0 to 1.0 by .10) label=(angle=90);
proc gplot data=outdata;
plot S*time /vaxis=axis1;
symbol1 i=join c=black line=1;
symbol2 i=join c=black line=2;
run;
```

<table>
<thead>
<tr>
<th>Obs</th>
<th>time</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0000</td>
<td>1.00000</td>
</tr>
<tr>
<td>2</td>
<td>0.4600</td>
<td>0.97872</td>
</tr>
<tr>
<td>3</td>
<td>1.1170</td>
<td>0.95745</td>
</tr>
<tr>
<td>4</td>
<td>4.7639</td>
<td>0.93617</td>
</tr>
<tr>
<td>5</td>
<td>5.7495</td>
<td>0.91489</td>
</tr>
<tr>
<td>6</td>
<td>6.0000</td>
<td>0.87234</td>
</tr>
<tr>
<td>7</td>
<td>13.5359</td>
<td>0.85106</td>
</tr>
<tr>
<td>8</td>
<td>13.8973</td>
<td>0.82979</td>
</tr>
<tr>
<td>9</td>
<td>22.7351</td>
<td>0.80608</td>
</tr>
</tbody>
</table>

12. You can also use this technique to get predicted curves for individuals at particular values of the predictors. For example, if you want to get the predicted survival function for a woman who has a low calcium intake and low bone mineral density:

```sas
data mycovs;
input calc1 bmc1;
datalines;
800 1900
;
run;
```

Similar to K-M curve, but smooth!
**STATS 210 SAS LAB TEN, August 9, 2004**

```sas
proc phreg data=stats210.runners;
  model time*sf1(0)=calc1 bmc1 / risklimits;
  baseline out=outdata covariates=mycovs survival=S;
run;
proc print data=outdata;
run;
proc gplot data=outdata;
  plot S*time=calc1 /vaxis=axis1 nolegend;
  symbol1 i=join c=black line=1;
  symbol2 i=join c=black line=2;
run;
```

Then tell SAS to use those covariates to calculate the estimated survival curve.

As calc1 and bmc1 distinguish the two sets of survival values (See below), I can use either one to ask for separate survival plots for a woman with the covariates I’ve specified vs. a woman with the mean values of the covariates (which SAS automatically calculates).

---

**Survival Curve**

- **For woman with calcium intake of 800 mg/day and bone mineral content at baseline of 1900 g.**

<table>
<thead>
<tr>
<th>Obs</th>
<th>calc1</th>
<th>bmc1</th>
<th>time</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>800.00</td>
<td>1900.00</td>
<td>0.0000</td>
<td>1.00000</td>
</tr>
<tr>
<td>2</td>
<td>800.00</td>
<td>1900.00</td>
<td>0.4600</td>
<td>0.95506</td>
</tr>
<tr>
<td>3</td>
<td>800.00</td>
<td>1900.00</td>
<td>1.1170</td>
<td>0.91096</td>
</tr>
<tr>
<td>4</td>
<td>800.00</td>
<td>1900.00</td>
<td>4.7639</td>
<td>0.86738</td>
</tr>
<tr>
<td>5</td>
<td>800.00</td>
<td>1900.00</td>
<td>5.7495</td>
<td>0.82431</td>
</tr>
<tr>
<td>6</td>
<td>800.00</td>
<td>1900.00</td>
<td>6.0000</td>
<td>0.73823</td>
</tr>
<tr>
<td>7</td>
<td>800.00</td>
<td>1900.00</td>
<td>13.5359</td>
<td>0.69563</td>
</tr>
<tr>
<td>8</td>
<td>800.00</td>
<td>1900.00</td>
<td>13.8973</td>
<td>0.65431</td>
</tr>
<tr>
<td>9</td>
<td>800.00</td>
<td>1900.00</td>
<td>22.7351</td>
<td>0.60964</td>
</tr>
<tr>
<td>10</td>
<td>1389.34</td>
<td>2199.90</td>
<td>0.0000</td>
<td>1.00000</td>
</tr>
<tr>
<td>11</td>
<td>1389.34</td>
<td>2199.90</td>
<td>0.4600</td>
<td>0.98445</td>
</tr>
<tr>
<td>12</td>
<td>1389.34</td>
<td>2199.90</td>
<td>1.1170</td>
<td>0.96871</td>
</tr>
<tr>
<td>13</td>
<td>1389.34</td>
<td>2199.90</td>
<td>4.7639</td>
<td>0.95266</td>
</tr>
<tr>
<td>14</td>
<td>1389.34</td>
<td>2199.90</td>
<td>5.7495</td>
<td>0.93627</td>
</tr>
<tr>
<td>15</td>
<td>1389.34</td>
<td>2199.90</td>
<td>6.0000</td>
<td>0.90173</td>
</tr>
<tr>
<td>16</td>
<td>1389.34</td>
<td>2199.90</td>
<td>13.5359</td>
<td>0.88364</td>
</tr>
<tr>
<td>17</td>
<td>1389.34</td>
<td>2199.90</td>
<td>13.8973</td>
<td>0.86539</td>
</tr>
<tr>
<td>18</td>
<td>1389.34</td>
<td>2199.90</td>
<td>22.7351</td>
<td>0.84478</td>
</tr>
</tbody>
</table>

- **For woman with the sample mean values of calcium and bmc.**

Define the survival curve for a woman with the sample mean values of calcium and bone mineral content at baseline.

Use the option “/ nmean” in the baseline statement of PROC PHREG to suppress estimation of survival function at the mean values of the covariates.
13. We can also use these plots to assess the validity of the proportional hazards assumption.

```
axis1 label=(angle=90);
proc phreg data=stats210.runners;
    model time*sf1(0)=;
    strata stressf;
    baseline out=outdata loglogs=lls;
run;
```

```
proc gplot data=outdata;
   title 'Evaluate proportional hazards assumption for variable: stressf';
   plot lls*time=stressf /vaxis=axis1;
   symbol1 i=join c=black line=1;
   symbol2 i=join c=black line=2;
run;
```

Note that stressf has been removed from the model statement; stratifying by stressf allows SAS to assume different baseline hazards for each stressf group (which can later be compared to test PH assumption)...

Similar plot can also be made in LIFETEST by asking for the lls plot:
```
proc lifetest data=stats210.runners
   plots=(lls) graphics
censoredsymbol=none;
    time time*sf1(0);
    strata stressf;
run;
```