Lab Nine: Kaplan-Meier methods for survival analysis

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

1. Finish up diagnostics from logistic regression.
2. Understand the structure of survival analysis data: time and censor variables.
3. Practice enhanced graphs using PROC GPLOT.
   a. Use the TITLE, SYMBOL, LABEL, AXIS, and LEGEND statements (which are global statements).
   b. Use different symbols for different values of a classification variable (such as censored/failed).
   c. Directly edit and enhance graphs using point and click features.
   d. Export graphs as image files.
4. Know where to go for help on SAS/GRAPH: http://v8doc.sas.com/sashtml/
5. Use PROC LIFETEST to generate Kaplan-Meier product-limit survival estimates, and understand the output of the LIFETEST procedure.
6. Produce a simple Kaplan-Meier curve.
7. Generate confidence limits for the Kaplan-Meier curve.
8. Use the LIFETEST procedure to compare survival times of two or more groups.

Recommended reading in Walker: Chapter 21
LAB EXERCISE STEPS:

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. Finish the last bit from lab eight:
   As with linear regression, it is important to check diagnostics. For example, how far off are our predicted risks (a probability of fracture) from what actually happened (1 or 0)? Refer to the lab on linear regression for SAS code to plot residuals against predictors.

   ```sas
   proc logistic data = stats210.runners;
      class quartneck (ref="0") ;
      model sf1 (event="1") = quartneck /iplots;
      output out = outdata l = Lower p = Predicted u = Upper resdev=residuals;
   run;
   run;
   proc print data=outdata;
      var quartneck lower predicted upper residuals;
   run;
   ```
Many different types of residuals are different versions of observed-expected.

**Predicted probability of a fracture**

### FROM PROC PRINT:

<table>
<thead>
<tr>
<th>Obs</th>
<th>Lower</th>
<th>Predicted</th>
<th>Upper</th>
<th>residuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.090485</td>
<td>0.27273</td>
<td>0.58566</td>
<td>1.61201</td>
</tr>
<tr>
<td>2</td>
<td>0.090485</td>
<td>0.27273</td>
<td>0.58566</td>
<td>1.61201</td>
</tr>
<tr>
<td>3</td>
<td>0.090485</td>
<td>0.27273</td>
<td>0.58566</td>
<td>-0.79806</td>
</tr>
<tr>
<td>4</td>
<td>0.090485</td>
<td>0.27273</td>
<td>0.58566</td>
<td>-0.79806</td>
</tr>
<tr>
<td>18</td>
<td>0.041982</td>
<td>0.16667</td>
<td>0.47720</td>
<td>-0.60386</td>
</tr>
<tr>
<td>19</td>
<td>0.041982</td>
<td>0.16667</td>
<td>0.47720</td>
<td>-0.60386</td>
</tr>
<tr>
<td>20</td>
<td>0.041982</td>
<td>0.16667</td>
<td>0.47720</td>
<td>-0.60386</td>
</tr>
<tr>
<td>21</td>
<td>0.041982</td>
<td>0.16667</td>
<td>0.47720</td>
<td>-0.60386</td>
</tr>
<tr>
<td>22</td>
<td>0.041982</td>
<td>0.16667</td>
<td>0.47720</td>
<td>-0.60386</td>
</tr>
<tr>
<td>23</td>
<td>0.041982</td>
<td>0.16667</td>
<td>0.47720</td>
<td>-0.60386</td>
</tr>
<tr>
<td>24</td>
<td>0.011606</td>
<td>0.08335</td>
<td>0.41320</td>
<td>-0.41721</td>
</tr>
<tr>
<td>25</td>
<td>0.011606</td>
<td>0.08335</td>
<td>0.41320</td>
<td>-0.41721</td>
</tr>
<tr>
<td>31</td>
<td>0.011606</td>
<td>0.08335</td>
<td>0.41320</td>
<td>-0.41721</td>
</tr>
<tr>
<td>38</td>
<td>0.082773</td>
<td>0.25000</td>
<td>0.55182</td>
<td>-0.75853</td>
</tr>
<tr>
<td>39</td>
<td>0.082773</td>
<td>0.25000</td>
<td>0.55182</td>
<td>-0.75853</td>
</tr>
</tbody>
</table>

### OUTPUT from 'iplots' option:

The LOGISTIC Procedure

Outlier? Observation 26 had a fracture but was in the second (lowest risk) quartile of bone density.
SURVIVAL ANALYSIS

4. To save time, I have already created a variable “time” which is either the interval between DXA 1 and DXA 3 (total time in the study) or is the time from entry into the study until first stress fracture.

Plot daily calcium intake (mg) at baseline against survival time (time to fracture or last observation) to see if calcium predicts fracture-free survival time:

```sas
goptions reset=all; *resets graphing options;
proc gplot data=stats210.runners;
title1 'Plot of survival time (months) vs. daily calcium intake';
plot time*calc1 /
vaxis = 0 6 12 18 24 30
haxis = 0 500 1000 1500 2000 2500 3000;
run; quit;
```

5. Make the graph a little fancier by adding the following features: change the plotting symbol color, shape, and size; reduce minor tick marks to 1 tick between every major tick; and change 'calc1' label on the x-axis to 'Daily calcium (mg)'. Refer to the GPLOT appendix for more plotting symbol options. To save time, just add the underlined elements to the previously entered code.

```sas
symbol1 va
proc gplot data=stats210.runners;
   title1 'Plot of survival time (months) vs. daily calcium intake';
   label calc1='Daily calcium (mg)';
   plot time*calc1 /
vaxis = 0 6 12 18 24 30 vminor=1
haxis = 0 500 1000 1500 2000 2500 3000 hminor=1;
run; quit;
```
6. Make the graph even fancier by differentiating between those participants who fractured and those who were censored (add sf1 as a classification variable and assign different plotting symbols for fractured and censored). Also add the following features: use global statements to specify the axes, including turning the x-axis label 90 degrees and changing font size and type; call these axes later within PROC GPLOT. As we are using a classification variable, SAS will automatically add a legend; format variable sf1 for appearance in the legend. For symbol 2, try a different plotting symbol using the appendix (I’ve chosen a shamrock).

```sas
proc format;
  value sf
    1="fracture"
    0="no fracture";
run;
goptions reset=all;
axis1 order= (0 to 30 by 6) label=(height=4pct font='Times New Roman' angle=90);
axis2 order= (0 to 4000 by 500) label=(height=4pct font='Times New Roman');
symbol1 v=circle c=blue h=1 w=1;
symbol2 value=& color=red h=1 w=1;
proc gplot data=stats210.runners;
title1 'Plot of survival time (months) vs. calcium (mg) by stress fracture (yes/no)';
  label time='Survival Time (Months)';
  label calc1='Daily calcium (mg)';
  label sf1='Stress fracture: ';
  format sf1 sf.;
  plot time*calc1=sf1 / vaxis = axis1 haxis=axis2 vminor=1 hminor=1;
run; quit;
```

Plot of survival time (months) vs. age by stress fracture (yes/no)
7. Finally, make the legend fancier by specifying where to put the legend (bottom, center, outside) and giving it a frame with light gray (ligr) background and black border.

```sas
legend1 label=(\textit{Stress Fracture}) shape=symbol(1,2) frame cframe=ligr cborder=black position=(bottom center outside);

proc gplot data=stats210.runners;
  title1 'Plot of survival time (months) vs. calcium by stress fracture (yes/no)';
  label time='Survival Time (Months)';
  label calc1='Daily calcium (mg)';
  label sf1='Stress fracture:,'
  format sf1 sf.;
  plot time*calc1=sf1 /
    vaxis = axis1 haxis=axis2 vminor=1 hminor=1 legend=legend1;
run; quit;
```
8. Plot the Kaplan-Meier curve for stress fracture vs. time-to-event:

The convention for all survival analyses in SAS is:
time*censor(censor value), where time is the time until event or censoring, and censor is a binary indicator variable that tells whether an individual had the event or was censored. Give SAS the value that represents censored in the parentheses.

/*Plot KM curve*/
goptions reset=all;
proc lifetest data=stats210.runners plots=(s) graphics censoredsymbol=none;
time time*sf1(0);
title 'Kaplan-Meier plot of Fracture-free survivorship';
symbol v=none;
run;

The steeper drop at 6 months into the study indicates that more than one fracture occurred at 6 months; in fact, when fracture dates were unknown, we gave a best guess of 6 months for the time to fracture (if the fracture occurred sometime during the first year of follow-up), so this is an artifact of how we dealt with missing data rather than a real phenomenon.

Indicates that cumulative probability of fracture is about 20%.

Notice that the curve plateaus after 24 months; that’s because participants were supposed to finish the study in 2 years. The flat line from 24 to 48 months represents a single participant who took much too long to finish and never had a fracture. When you see a flat curve at the end like this, it usually indicates that there are very few participants remaining in the study or remaining at risk, and does not represent a reduction in risk of the event.
## Kaplan-Meier Plot of Fracture-free Survivorship

The Kaplan-Meier plot of fracture-free survivorship is shown below. The table provides product-limit survival estimates at various time points, along with the number of failures, number of censored observations, and the estimated survival to each event time. The last of the tied cases, when ties exist, is reported. Censored observations are starred.

### Summary Statistics for Time Variable time

#### Quartile Estimates

<table>
<thead>
<tr>
<th>Percent</th>
<th>Estimate</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>13.5359</td>
<td>(Lower, Upper)</td>
</tr>
<tr>
<td>50</td>
<td>.</td>
<td>19.9616, 75</td>
</tr>
</tbody>
</table>

The estimated median death time and 95% confidence interval is not estimable here. Estimated mean fracture-free “survival” time is 19.9616 with a standard error of 0.9930.
9. Next, we would like to compare fracture rates between treatment and control groups. Use the `strata` statement in PROC LIFETEST.

```sas
proc format;
  value treat =
    0 = "control";
run;
proc lifetest data=stats210.runners plots=(s) graphics event=none censored=none maxtime=24;
  time time*sf1(0);
  strata treatr;
  format treatr treat.;
  title 'Kaplan-Meier plot of fracture-free survivorship by treatment';
run;
```

Formatting improves the appearance of the legend that appears on the Kaplan-Meier graph.

I've added the option “maxtime=24” to cut the graph at 24 months of follow-up.

10. To change the color and appearance of the lines for black-and-white printouts, add the following two lines to your above code (between the title statement and run) and rerun:

```sas
symbol1 v=none c=black w=2 i=join line=1;
symbol2 v=none c=black w=2 i=join line=2;
```

**DUE FROM LAB NINE:** Use right click ➔ Edit ➔ Edit current graph to get to the editing screen for this graph from step 10. Use the font tool to add your name anywhere on the graph. Exit the editing screen and use right click ➔ File ➔ Export as image to save your graph. Email me the graph.
Log-rank test of the null hypothesis of no difference in survival experience between controls and treatment reveals little difference between the groups (p=.889, log-rank test):

Test of Equality over Strata

<table>
<thead>
<tr>
<th>Test</th>
<th>Chi-Square</th>
<th>DF</th>
<th>Pr &gt; Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log-Rank</td>
<td>0.0195</td>
<td>1</td>
<td>0.8890</td>
</tr>
<tr>
<td>Wilcoxon</td>
<td>0.0063</td>
<td>1</td>
<td>0.9370</td>
</tr>
<tr>
<td>-2Log(LR)</td>
<td>0.0042</td>
<td>1</td>
<td>0.9482</td>
</tr>
</tbody>
</table>

11. 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) by changing strata statement (and title) as below (cut and paste from above and change the strata statement).

```
proc lifetest data=stats210.runners plots=(s) graphics censoredsymbol=none maxtime=24;
  time time*sf1(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: (-∞,800) [800, 1500) [1500, ∞). 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.
12. Request log-survival plot and log-log survival plots for the Kaplan-Meier curve with stressf (stress fracture prior to baseline) as the groups.

```
proc lifetest data=stats210.runners plots=(ls, lls, s) graphics
censoredsymbol=none maxtime=24;
  time time*sf1(0);      
  strata stressf;        
  title 'Kaplan-Meier plot of fracture-free survivorship by prior fracture';
  symbol1 v=none c=black w=2 i=join line=1; 
  symbol2 v=none c=black w=2 i=join line=2; 
  symbol2 v=none c=black w=2 i=join line=3;
run;
```

The plot tells us how the hazard changes with time:
- For example, if the hazard is constant (no change over time), should be a straight line with origin at 0.
- Evaluates proportional hazards assumption:
  Should be roughly parallel curves if the risks are proportional.

Equivalent to the cumulative hazard function; if you're mathematically inclined:

\[
- \log S(t) = \int_0^t h(u)du
\]

The plot tells us how the hazard changes with time:
- For example, if the hazard is constant (no change over time), should be a straight line with origin at 0.