Applied Survival Analysis
Lab 5: More on Cox Proportional Hazards Model

Today, we are going to see how to construct confidence intervals and tests for hazard ratios. Also, we are going to compare nested models using likelihood ratio tests. Then we are going to learn how to estimate the baseline survival function, predicted medians and P-year survival.

1. *C.I., Wald test and Likelihood Ratio test: MAC Dataset*

This time we are interested in the time to MAC disease and not in time to death. So we are going to *stset* the data in the following way:

`stset mactime, failure(macstat)`

```
     failure event:  macstat ~= 0 & macstat ~= .
obs. time interval:  (0, mactime]
 exit on or before:  failure

------------------------------------------------------------------------------
     1177  total obs.
       26  obs. end on or before enter()
------------------------------------------------------------------------------
     1151  obs. remaining, representing
      121  failures in single record/single failure data
   489509  total analysis time at risk, at risk from t =         0
                             earliest observed entry t =         0
                                  last observed exit t =       827
```

First we are going to fit the following model:

Model 1: $\lambda(t, X) = \lambda_0(t)\exp(\beta_1 KARNOF + \beta_2 RIF + \beta_3 CLARI)$

`stcox  karnof rif clari, nohr`

```
         failure _d:  macstat
   analysis time _t:  mactime

(iterations )
Refining estimates:
Iteration 0:   log likelihood = -754.52813

Cox regression -- Breslow method for ties

No. of subjects =         1151                     Number of obs   =      1151
No. of failures =          121
Time at risk    =       489509
                                                   LR chi2(3)      =     32.01
Log likelihood  =   -754.52813                     Prob > chi2     =    0.0000

------------------------------------------------------------------------------
      _t |
      _d |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  karnof | -.0448295   .0106355      -4.215   0.000      -.0656747   -.0239843
     rif |  .8723819   .2369497       3.682   0.000       .4079691    1.336795
   clari |  .2760775   .2580215       1.070   0.285      -.2296354    .7817903
------------------------------------------------------------------------------
```

(a) What is the hazard ratio of the Karnofsky score status? What is the interpretation of this hazard ratio?

(b) Using $[L,U] = [e^{\hat{\beta}-1.96se(\hat{\beta})}, e^{\hat{\beta}+1.96se(\hat{\beta})}]$. Construct the 95% confidence interval of the estimated hazard ratio in (a), interpret your result.

(c) Test the effect of the Karnofsky score using Wald test. State your null and alternative hypothesis. What do you conclude?

Next we want to add the effect of CD4, so we need to fit the following model:

Model 2: $\lambda(t,X) = \lambda_0(t)\exp(\beta_1 KARNOF + \beta_2 RIF + \beta_3 CLARI + \beta_4 CD4)$

`stcox  karnof rif clari cd4, nohr`

```
         failure _d:  macstat
   analysis time _t:  mactime

Iteration 0:   log likelihood = -770.53218
Iteration 1:   log likelihood = -740.59073
Iteration 2:   log likelihood = -738.68473
Iteration 3:   log likelihood = -738.66226
Iteration 4:   log likelihood = -738.66225
Refining estimates:
Iteration 0:   log likelihood = -738.66225

Cox regression -- Breslow method for ties

No. of subjects =         1151                     Number of obs   =      1151
No. of failures =          121
Time at risk    =       489509
                                                   LR chi2(4)      =     63.74
Log likelihood  =   -738.66225                     Prob > chi2     =    0.0000

------------------------------------------------------------------------------
      _t |
      _d |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  karnof |  -.0368538   .0106652     -3.456   0.001      -.0577572   -.0159503
     rif |    .880338   .2371111      3.713   0.000       .4156089    1.345067
   clari |   .2530205   .2583478      0.979   0.327      -.253332     .7593729
     cd4 |  -.0183553   .0036839     -4.983   0.000      -.0255757   -.0111349
------------------------------------------------------------------------------
```

To construct a Likelihood Ratio test comparing this model (saturated) to model 1 (reduced) in STATA, you use the **`lrtest`** command. But first you have to fit the saturated (bigger) model, save it and then fit the smaller model to get the right likelihood ratio test in STATA. So after model 2 we would fit again model 1; the sets of commands are the following:

`stcox  karnof rif clari cd4` (**Model 2**)

`estimates store B`
(specifies that the summary statistics associated with the most recently estimated model are to be saved as name. The saturated model is typically saved by typing "`estimates` (or just `est`) `store B`".)

`stcox  karnof rif clari` (**Model 1**)
`est store A`

`lrtest A B`

```
Cox:  likelihood-ratio test                          chi2(1)     =      31.73
(Assumption: A nested in B)                          Prob > chi2 =     0.0000
```

(d) Compute the likelihood ratio test by hand and confirm that you get the same result as above. What do you conclude from this result?

To conduct an overall test of treatment effect we can use the **`test`** command in STATA:

`stcox  karnof rif clari cd4` (Fit model 2 first to test the treatment effect in this model).

`test rif clari`

```
 ( 1)  rif = 0.0
 ( 2)  clari = 0.0

           chi2(  2) =   17.01
         Prob > chi2 =    0.0002
```

The **`test`** command can also be used to test whether there is a difference between the **`rif`** and **`clari`** treatment arms:

`test rif=clari`

```
 ( 1)  rif - clari = 0.0

           chi2(  1) =    8.76
         Prob > chi2 =    0.0031
```

2. *Survival Function, Predicted Medians and P-year Survival: Nursing Home Data*

(Morris et al., *Case Studies in Biometry*, Ch 12)

We are going to consider the same example as last time (*nurshome.dta*).

Again before starting any analysis we have to *stset* our data:

`stset los, failure(fail)`

To predict the baseline survival we use the option **`basesurv`** after the **`stcox`** command:

`stcox married health, basesurv(prsurv)` (Name baseline survival *prsurv*)

```
         failure _d:  fail
   analysis time _t:  los

Iteration 0:   log likelihood = -8556.5713
Iteration 1:   log likelihood = -8534.0911
Iteration 2:   log likelihood = -8533.9783
Iteration 3:   log likelihood = -8533.9783
Refining estimates:
Iteration 0:   log likelihood = -8533.9783

Cox regression -- Breslow method for ties

No. of subjects =         1591                    Number of obs   =      1591
No. of failures =         1269
Time at risk    =       386211
                                                  LR chi2(2)      =     45.19
Log likelihood  =   -8533.9783                    Prob > chi2     =    0.0000

------------------------------------------------------------------------------
      _t |
      _d | Haz. Ratio   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
 married |   1.345394   .0971282      4.110   0.000       1.167881    1.549889
  health |    1.17993   .0368631      5.296   0.000       1.109847    1.254438
------------------------------------------------------------------------------
```

`sort los`

`list  los prsurv in 1/10`

```
           los      prsurv
  1.         1   .99252899
  2.         1   .99252899
  3.         1   .99252899
  4.         1   .99252899
  5.         1   .99252899
  6.         1   .99252899
  7.         1   .99252899
  8.         1   .99252899
  9.         1   .99252899
 10.         1   .99252899
```

To get the predicted survival for subgroups we will use the following set of commands:

predict betaz, xb (**xb** calculates the linear prediction from the estimated model)

gen newterm=exp(betaz)

gen predsurv=prsurv^newterm $(S_i(t) = [S_0(t)]^{\exp(\beta Z_i)})$

sort married health los

list married health los predsurv in 1/20

```
          married       health        los   predsurv
  1. Not Married  Second Best          1   .9896138
  2. Not Married  Second Best          1   .9896138
  3. Not Married  Second Best          1   .9896138
  4. Not Married  Second Best          1   .9896138
  5. Not Married  Second Best          1   .9896138
  6. Not Married  Second Best          1   .9896138
  7. Not Married  Second Best          1   .9896138
  8. Not Married  Second Best          2    .981557
  9. Not Married  Second Best          2    .981557
 10. Not Married  Second Best          2    .981557
 11. Not Married  Second Best          3   .9772769
 12. Not Married  Second Best          3   .9772769
 13. Not Married  Second Best          4   .9691724
 14. Not Married  Second Best          4   .9691724
 15. Not Married  Second Best          4   .9691724
 16. Not Married  Second Best          5   .9586483
 17. Not Married  Second Best          6    .951448
 18. Not Married  Second Best          6    .951448
 19. Not Married  Second Best          7   .9427774
 20. Not Married  Second Best          8   .9360114
```

Next we are going to create the four groups of interest (single+healthy, single+unhealthy, married+healthy and married+unhealthy) :

gen group=1 if married==0 & health==2
(1292 missing values generated)

replace group=2 if married==0 & health==5
(135 real changes made)

replace group=3 if married==1 & health==2
(42 real changes made)

replace group=4 if married==1 & health==5
(33 real changes made)

Then generate the predicted survival for these subgroups:

gen predsur1=predsurv if group==1
(1292 missing values generated)

gen predsur2=predsurv if group==2
(1456 missing values generated)

gen predsur3=predsurv if group==3
(1549 missing values generated)

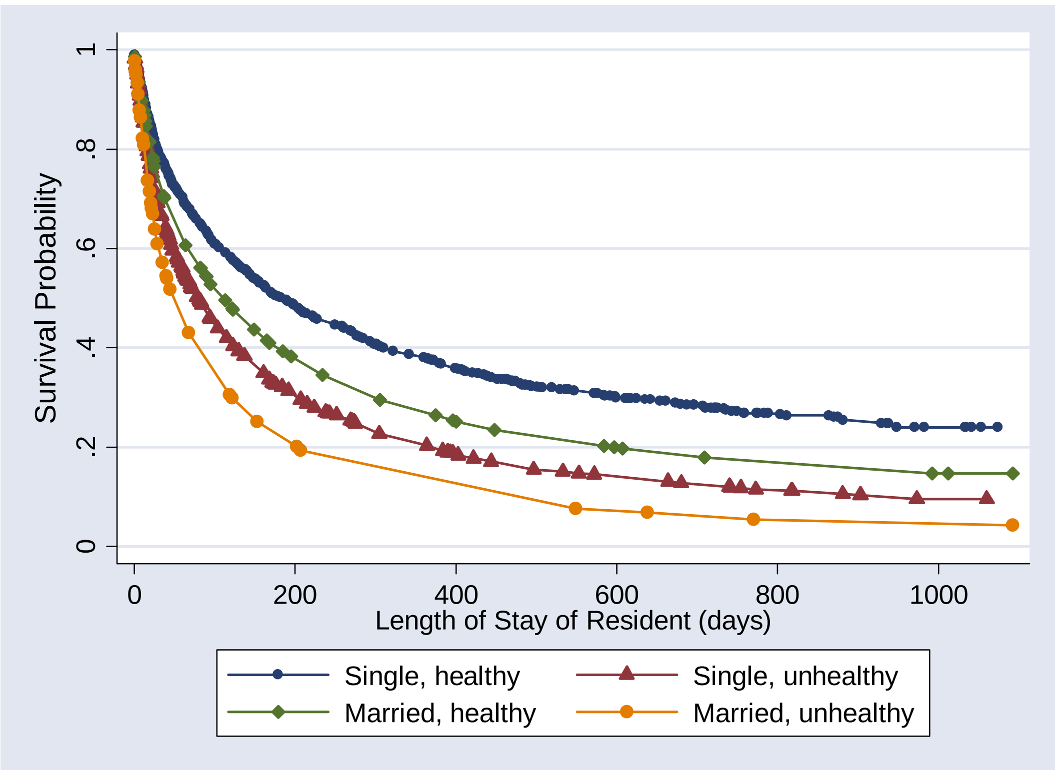
gen predsur4=predsurv if group==4
(1558 missing values generated)

And label the predicted survivals:

```
lab var  predsur1"Single, healthy"

lab var  predsur2"Single, unhealthy"

lab var  predsur3"Married, healthy"

lab var  predsur4 "Married, unhealthy"
```

If we want to get a visual picture of what the proportional hazards assumption implies for these four subgroups we can use the following command:

```
sort los

scatter predsur1  predsur2  predsur3  predsur4 los, c(l l l l) s(o T
d O) l1(Survival Probability)
```



(e) Which subgroup has the longest length of stay?

To get the predicted medians we can use the following approaches:

Kaplan-Meier Approach:

stsum, by(group)

```
         failure _d:  fail
   analysis time _t:  los

         |               incidence       no. of    |------ Survival time -----|
group    | time at risk     rate        subjects        25%        50%        75%
---------+---------------------------------------------------------------------
 Single, |        81792   .0027753           299         43        151        654
 Single, |        23594   .0051284           135         18         62        240
Married, |         9751   .0035894            42         24         95        375
Married, |         4313   .0069557            33          8         23        119
---------+---------------------------------------------------------------------
   total |       119450   .0034575           509         27        100        412
```

Or we can list the predicted survivals of each group around 50% :

list married health los predsur1 if predsur1>0.49 & predsur1<0.51

```
          married       health        los   predsur1
1387. Not Married  Second Best        172   .5088611
1391. Not Married  Second Best        176   .5052575
1392. Not Married  Second Best        180   .5028501
1393. Not Married  Second Best        180   .5028501
1394. Not Married  Second Best        182   .5016459
1397. Not Married  Second Best        189   .4968184
1398. Not Married  Second Best        191    .494399
```

list married health los predsur2 if predsur2>0.49 & predsur2<0.51

```
          married       health        los   predsur2
1315. Not Married        Worst         78   .5026844
1316. Not Married        Worst         81   .4971449
1317. Not Married        Worst         82   .4943793
1322. Not Married        Worst         83   .4923071
```

list married health los predsur3 if predsur3>0.49 & predsur3<0.51

```
          married       health        los   predsur3
1342.     Married  Second Best        113   .4953526
```

list married health los predsur4 if predsur4>0.43 & predsur4<0.51

```
          married       health        los   predsur4
1300.     Married        Worst         68   .4300353
```