Session 7: Parametric survival analysis

To generate parametric survival analyses in SAS we use `PROC LIFEREG`. For exponential regression analysis of the nursing home data the syntax is as follows:

```
data nurshome;
     infile 'nurshome.dat';
     input los age rx gender married health fail;
     label los='Length of stay'
           rx='Treatment'
           married='Marriage status'
           health='Health index'
           fail='Censoring index';
     format married marfmt.;
run;
```

```
proc lifereg data=nurshome outest=expoutest;
     model los*fail(0)=gender/dist=exponential;
     title 'Exponential regression for the nursing home data';
     output out=expsurv xbeta=exp_xb;
run;
```

The output is as follows:

```
                          Exponential regression for the nursing home data

                                       The LIFEREG Procedure

                                         Model Information

               Data Set                      WORK.NURSHOME
               Dependent Variable                 Log(los)    Length of stay
               Censoring Variable                     fail    Censoring index
               Censoring Value(s)                        0
               Number of Observations                 1591
               Noncensored Values                     1269
               Right Censored Values                   322
               Left Censored Values                      0
               Interval Censored Values                  0
               Name of Distribution            Exponential
               Log Likelihood                 -3320.476626


                              Number of Observations Read        1591
                              Number of Observations Used        1591


                                          Fit Statistics

                          -2 Log Likelihood                   6640.953
                          AIC (smaller is better)             6644.953
                          AICC (smaller is better)            6644.961
                          BIC (smaller is better)             6655.697


            Algorithm converged.
```

```
                     Type III Analysis of Effects

                                      Wald
              Effect        DF     Chi-Square     Pr > ChiSq

              gender         1        69.5062         <.0001


            Analysis of Maximum Likelihood Parameter Estimates

                            Standard   95% Confidence     Chi-
Parameter      DF Estimate     Error       Limits        Square Pr > ChiSq

Intercept       1   5.8421    0.0333   5.7769   5.9074 30785.8      <.0001
gender          1  -0.5162    0.0619  -0.6375  -0.3948   69.51      <.0001
Scale           0   1.0000    0.0000   1.0000   1.0000
Weibull Shape   0   1.0000    0.0000   1.0000   1.0000


                         The LIFEREG Procedure

                     Lagrange Multiplier Statistics

                 Parameter      Chi-Square     Pr > ChiSq

                 Scale            337.5980         <.0001
```

The estimate of the coefficient associated with gender is β=-0.5162 corresponding to a reduction in the hazard for discharge from the nursing home among men (HR=exp(-0.5162)=0.597).

The weibull analysis of the same data set is done as follows:

```
proc lifereg data=nurshome outest=weiboutest;
      model los*fail(0)=gender/dist=weibull;
      title 'Weibull regression for the nursing home data';
      output out=weibsurv xbeta=weib_xb;
run;
```

Producing the following output:

```
              Weibull regression for the nursing home data

                         The LIFEREG Procedure

                           Model Information

     Data Set                      WORK.NURSHOME
     Dependent Variable                 Log(los)    Length of stay
     Censoring Variable                     fail    Censoring index
     Censoring Value(s)                        0
     Number of Observations                 1591
     Noncensored Values                     1269
     Right Censored Values                   322
     Left Censored Values                      0
     Interval Censored Values                  0
     Name of Distribution                Weibull
     Log Likelihood                 -3045.276811


                Number of Observations Read        1591
                Number of Observations Used        1591
```

```
                              Fit Statistics

          -2 Log Likelihood                       6090.554
          AIC (smaller is better)                 6096.554
          AICC (smaller is better)                6096.569
          BIC (smaller is better)                 6112.670


  Algorithm converged.


                       Type III Analysis of Effects

                                          Wald
               Effect        DF    Chi-Square    Pr > ChiSq

               gender         1       44.4042        <.0001


            Analysis of Maximum Likelihood Parameter Estimates

                                 Standard   95% Confidence     Chi-
   Parameter      DF Estimate     Error        Limits        Square Pr > ChiSq

   Intercept       1    5.7564    0.0542    5.6502    5.8627 11280.0     <.0001
   gender          1   -0.6735    0.1011   -0.8716   -0.4754   44.40     <.0001
   Scale           1    1.6275    0.0378    1.5551    1.7032
   Weibull Shape   1    0.6144    0.0143    0.5871    0.6430
```

Note that the Weibull shape parameter is 0.6144 with 95% confidence interval (0.5871-0.6430), suggesting that the distribution is not exponential (i.e., that with 95% confidence the shape parameter is below 1.0).

To check the model we overlay the exponential and Weibull estimated survival probabilities with those produced by the Kaplan-Meier analysis below:

```
*Run a Kaplan-Meier analysis and produce the estimated survival;
proc lifetest data=nurshome outsurv=kmsurv;
     time los*fail(0);
       strata gender;
       title 'Kaplan Meier analysis of the nursing home data';
run;
```

As SAS cannot produce estimated survival probabilities from PROC LIFEREG (!) we run a macro provided by Paul Allison (you can find a version of this at http://www.ssc.upenn.edu/~allison/PREDICT.SAS).

The macro is as follows:

```
%macro predict (outest=, out=_last_,xbeta=,time=);
/********************************************************************
MACRO PREDICT produces predicted survival probabilities for specified
survival times, based on models fitted by LIFEREG. When fitting the
model with LIFEREG, you must request the OUTEST data set on the
PROC statement.  You must also request an OUTPUT data set with the
XBETA= keyword.

PREDICT has four parameters:

OUTEST is the name of the data set produced with the OUTEST option.
OUT is the name of the data set produced by the OUTPUT statement.
        Default is the last created data set.
XBETA is the name of the variable specified with the XBETA= keyword.
TIME is the specified survival time that is to be evaluated.

Example:  To get 5-year survival probabilities for every individual
in the sample (assuming that actual survival times are measured in
years);

%predict(outest=a, out=b, xbeta=lp, time=5).

Author:  Paul D. Allison, Univ. of Pennsylvania
         allison@ssc.upenn.edu
*********************************************************************/
data _pred_;
_p_=1;
set &outest  point=_p_;
set &out;
lp=&xbeta;
t=&time;
gamma=1/_scale_;
alpha=exp(-lp*gamma);
prob=0;
_dist_=upcase(_dist_);
if _dist_='WEIBULL' or _dist_='EXPONENTIAL' or _dist_='EXPONENT' then
prob=exp(-alpha*t**gamma);
if _dist_='LOGNORMAL' or _dist_='LNORMAL' then prob=1-probnorm((log(t)-
lp)/_scale_);
if _dist_='LLOGISTIC' or _dist_='LLOGISTC' then prob=1/(1+alpha*t**gamma);
if _dist_='GAMMA' then do;
  d=_shape1_;
  k=1/(d*d);
  u=(t*exp(-lp))**gamma;
  prob=1-probgam(k*u**d,k);
  if d lt 0 then prob=1-prob;
  end;
drop lp gamma alpha _dist_ _scale_ intercept
     _shape1_ _model_ _name_ _type_ _status_ _prob_ _lnlike_ d k u;
run;
proc print data=_pred_;
run;
%mend predict;
```

I have commented out the print procedure at the end to obviate unnecessary printing in the version uploaded to the website.

The macro and subsequent steps required are as follows:

```
* Run Paul Alison's macro PREDICT to predict survival function from
LIFEREG;
* Put it in the default directory and run it;
%include'allisons-predict-macro.sas';
```

```
proc print data=expsurv;
     title 'Generated data from exponential regression';
run;

proc print data=expoutest;run;

* Run Allison's macro:
OUTEST is the name of the data set produced with the OUTEST option,
       i.e., EXPOUTEST.
OUT is the name of the data set produced by the OUTPUT statement.
       Default is the last created data set. However, we explicitly put
EXPSURV.
XBETA is the name of the variable specified with the XBETA= keyword.
      This is exp_xb here.
TIME  is the specified survival time that is to be evaluated. We put
      LOS to get estimates for all LOS.;

%predict (outest=expoutest, out=expsurv, xbeta=exp_xb,time=los);

* Sort the data set for plotting;
proc sort data=_pred_ out=expsort;
     by gender los;
run;
```

And for Weibull,

```
proc print data=weibsurv;
     title 'Generated data from Weibull regression';
run;

proc print data=weiboutest;run;

* Run Allison's macro:
OUTEST is the name of the data set produced with the OUTEST option,
       i.e., WEIBOUTEST.
OUT is the name of the data set produced by the OUTPUT statement.
       Default is the last created data set. We put it in WEIBSURV.
XBETA  is the name of the variable specified with the XBETA= keyword.
       This is WEIB_xb here.
TIME   is the specified survival time that is to be evaluated. We put
       LOS to get estimates for all LOS.;

%predict (outest=weiboutest, out=weibsurv, xbeta=weib_xb,time=los);

* Sort the data set for plotting;
proc sort data=_pred_ out=weibsort;
     by gender los;
run;
```

```
*Now merge all three data sets;
data kmweibexp;
     merge kmsurv   (keep=los gender survival rename=(survival=kmsurv))
           weibsort (keep=los gender prob rename=(prob=weibsurv))
             expsort  (keep=los gender prob rename=(prob=expsurv));
     by gender los;
       *Add an observation to the exponential and Weibull survival
        estimates at LOS=0;
     if los=0 then do;
        weibsurv=1; expsurv=1;
     end;
     if gender=1 then do;
        kmsurv_1=kmsurv; expsurv_1=expsurv; weibsurv_1=weibsurv;
     end;else
     if gender=0 then do;
        kmsurv_0=kmsurv; expsurv_0=expsurv; weibsurv_0=weibsurv;
     end;
     drop kmsurv expsurv weibsurv;
run;

proc print data=kmweibexp;
     title 'All three data sets';
run;
```

Notice that we add a 1.0 to the estimated survival probability at LOS=0 (because this is generated by the Kaplan-Meier analysis but not PROC LIFEREG.

```
     if los=0 then do;
        weibsurv=1; expsurv=1;
     end;
```

Also notice that we merged

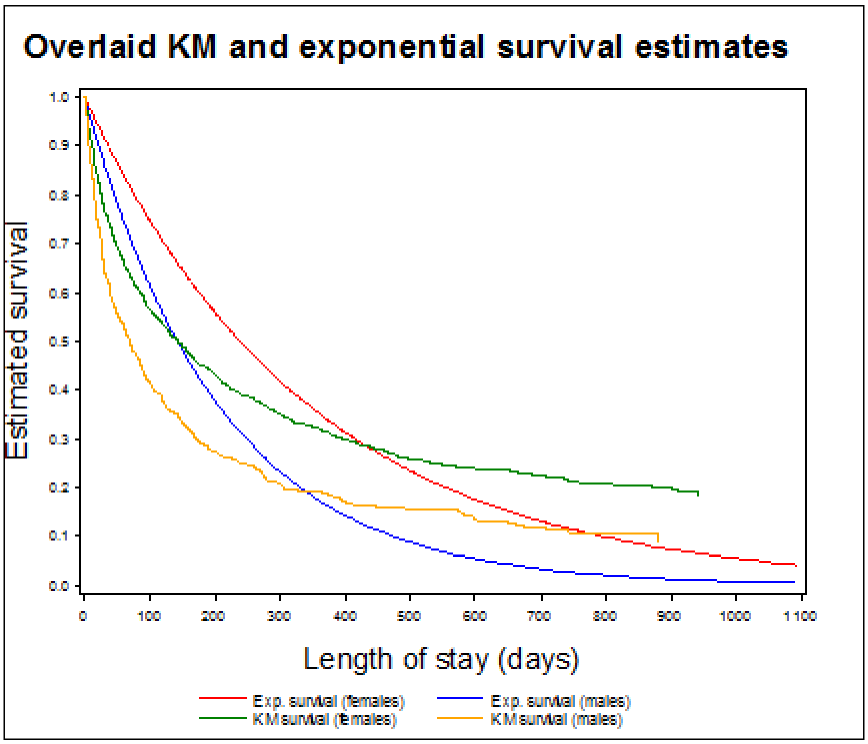
```
     by gender los;
```

This is critical! The printed results are as follows:

```
                                        All three data sets

                                             expsurv_    weibsurv_                expsurv_    weibsurv_
 Obs     gender     los     kmsurv_1            1            1        kmsurv_0        0            0

   1       0         0          .             .            .         1.00000      1.00000      1.00000
   2       0         1          .             .            .         0.98380      0.99710      0.97132
   3       0         1          .             .            .         0.98380      0.99710      0.97132
   4       0         1          .             .            .         0.98380      0.99710      0.97132
   5       0         1          .             .            .         0.98380      0.99710      0.97132
   .       .         .          .             .            .            .            .            .
   .       .         .          .             .            .            .            .            .
   .       .         .          .             .            .            .            .            .
1175       1            0    1.00000      1.00000      1.00000             .            .            .
1176       1            1    0.99282      0.99515      0.95694             .            .            .
1177       1            1    0.99282      0.99515      0.95694             .            .            .
1178       1            1    0.99282      0.99515      0.95694             .            .            .
1179       1            2    0.97608      0.99032      0.93483             .            .            .
1183       1            2    0.97608      0.99032      0.93483             .            .            .
1184       1            2    0.97608      0.99032      0.93483             .            .            .
   .       .         .          .             .            .            .            .            .
   .       .         .          .             .            .            .            .            .
   .       .         .          .             .            .            .            .            .
```

So we have created a data set sorted by gender and LOS and ready to plot. For the survival estimates generated by the exponential regression, this is done as follows:

```
* Overlay exponential and KM survival plots;
symbol1 i=join c=red;
symbol2 i=join c=blue;
symbol3 i=stepljs c=green;
symbol4 i=stepljs c=orange;
axis1 label=(h=2 a=90 'Estimated survival') minor=none;
axis2 label=(h=2 'Length of stay (days)') minor=none;
legend1 label=('') value=(h=1 'Exp. survival (females)' 'Exp. survival
(males)'
                              'KM survival (females)' 'KM survival
(males)');

*KM and exponential plot;
proc gplot data=kmweibexp;
     plot expsurv_0*los=1 expsurv_1*los=2
     kmsurv_0*los=3 kmsurv_1*los=4/overlay vaxis=axis1 haxis=axis2
                                 legend=legend1;
     title 'Overlaid KM and exponential survival estimates';
run;
```

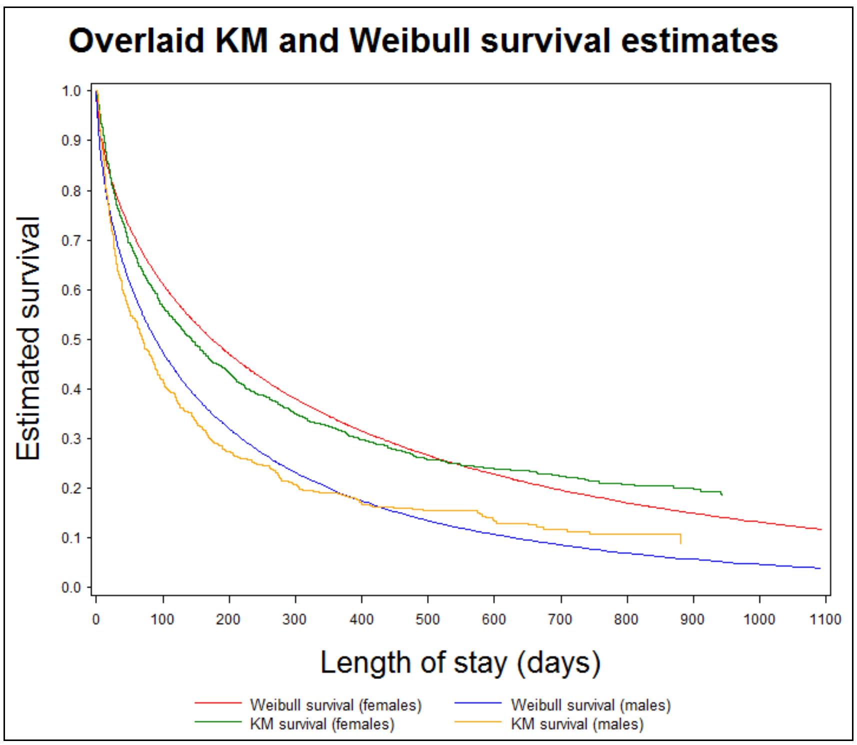


We note the poor agreement of the KM and exponential analyses. Given that the KM analysis makes no assumptions about the distribution of the survival times, this disagreement bodes badly for the validity of the exponential regression model.

We can do this for the Weibull as well:

```
*KM and Weibull plot;
legend2 label=('') across=2 value=(h=1 'Weibull survival (females)'
                                       'Weibull survival (males)'
                                       'KM survival (females)'
                                       'KM survival (males)');
proc gplot data=kmweibexp;
     plot weibsurv_0*los=1 weibsurv_1*los=2
          kmsurv_0*los=3 kmsurv_1*los=4/overlay vaxis=axis1 haxis=axis2
          legend=legend2;
     title 'Overlaid KM and Weibull survival estimates';
run;
```

Note that all we needed to change was the legend statement. The symbol definitions are carried over.



This is an improved fit but some issues remain.

Now let’s see what the Cox regression analysis looks like.

```
 *Cox proportional hazards analysis;
proc phreg data=nurshome ;
     model los*fail(0)=gender/ties=breslow;
     output out=coxphsurv survival=coxphsurv;
run;
```

The output is as follows:

```
                    Overlaid KM and Weibull survival estimates

                               The PHREG Procedure

                                Model Information

                  Data Set                 WORK.NURSHOME
                  Dependent Variable       los                Length of stay
                  Censoring Variable       fail               Censoring index
                  Censoring Value(s)       0
                  Ties Handling            BRESLOW


                           Number of Observations Read        1591
                           Number of Observations Used        1591


                      Summary of the Number of Event and Censored Values

                                                               Percent
                            Total       Event    Censored    Censored

                             1591        1269         322       20.24


                                Convergence Status

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


                              Model Fit Statistics

                                        Without           With
                           Criterion     Covariates     Covariates

                           -2 LOG L       17113.143      17075.121
                           AIC            17113.143      17077.121
                           SBC            17113.143      17082.267


                           Testing Global Null Hypothesis: BETA=0

                   Test                 Chi-Square       DF     Pr > ChiSq

                   Likelihood Ratio        38.0216        1         <.0001
                   Score                   40.8458        1         <.0001
                   Wald                    40.4091        1         <.0001


                               The PHREG Procedure

                    Analysis of Maximum Likelihood Estimates

                          Parameter      Standard                                    Hazard
     Parameter     DF      Estimate         Error    Chi-Square    Pr > ChiSq        Ratio

     gender         1       0.39473       0.06210       40.4091        <.0001        1.484
```

The results are very much like the Weibull analysis.

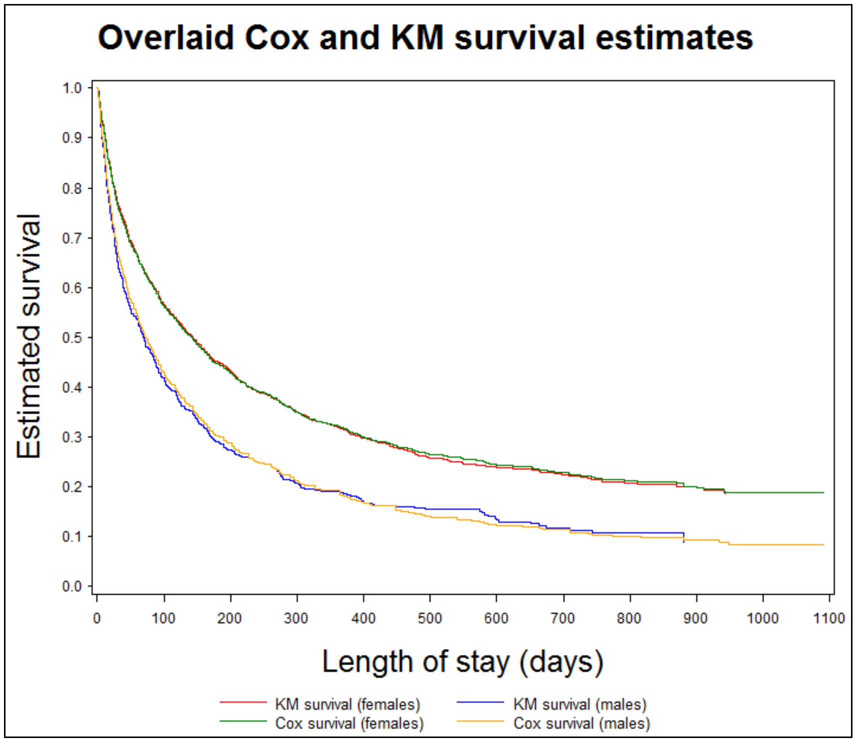
Now, let's merge the generated survival estimates from the Cox proportional hazards model.

```
proc sort data=coxphsurv out=coxphsort;
     by gender los;
run;

data kmweibexpcox;
     merge kmweibexp coxphsort(keep=gender los coxphsurv);
     by gender los;
     *Add an observation to the Cox PH survival estimates at LOS=0;
     if los=0 then coxphsurv=1;
     if gender=1 then coxphsurv_1=coxphsurv; else
     if gender=0 then coxphsurv_0=coxphsurv;
     drop coxphsurv;
run;
```

This is done as follows:

```
* Overlay Cox PH and KM survival plots;
symbol1 i=stepljs c=red;
symbol2 i=stepljs c=blue;
symbol3 i=stepljs c=green;
symbol4 i=stepljs c=orange;
axis1 label=(h=2 a=90 'Estimated survival') minor=none;
axis2 label=(h=2 'Length of stay (days)') minor=none;
legend1 label=('') value=(h=1 'KM survival (females)'
                         'KM survival (males)'
                         'Cox survival (females)'
                         'Cox survival (males)');

*Cox and Weibull plot;
proc gplot data=kmweibexpcox;
     plot KMsurv_0*los=1 KMsurv_1*los=2
     coxphsurv_0*los=3 coxphsurv_1*los=4/overlay vaxis=axis1 haxis=axis2
                                         legend=legend1;
     title 'Overlaid Cox and KM survival estimates';
run;
```

The resulting plot is as follows:



We see that the Cox proportional-hazards analysis produces a very good fit, which corresponds very well with the KM analysis.