Applied Survival Analysis
Lab 8: Parametric Survival Analysis

In today's lab, we are going to review the construction and interpretation of parametric models, including exponential and Weilbull models. Also we are going to evaluate how these models are fit using STATA. We will use the same example as in the lecture, the nursing home dataset (*nurshome.dta).*

Before we start the analysis we will transform the time to death into units of years:

```
gen losyr=los/365
lab var losyr "Length of Stay of Resident (years)"      ( To label )
```

The general command to get parametric models in STATA is with **`streg varlist , dist(distribution)`**. So if we would like to use the exponential model to test the gender variable, we would type the following command: (Note we have to ***stset*** our data first).

```
stset losyr, f(fail)

streg gender, dist(exponential)
```

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


Iteration 0:   log likelihood = -3352.5765
Iteration 1:   log likelihood =  -3321.966
Iteration 2:   log likelihood = -3320.4792
Iteration 3:   log likelihood = -3320.4766
Iteration 4:   log likelihood = -3320.4766

Exponential regression -- log relative-hazard form

No. of subjects =         1591                     Number of obs   =      1591
No. of failures =         1269
Time at risk    =  1058.112328
                                                   LR chi2(1)      =     64.20
Log likelihood  =   -3320.4766                     Prob > chi2     =    0.0000

------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.       z     P>|z|       [95% Conf. Interval]
-------------+----------------------------------------------------------------
      gender |   1.675625    .103746      8.337   0.000        1.48414    1.891815
------------------------------------------------------------------------------
```

STATA directly gives us the hazard ratio. To get the estimates of the coefficients we should add the option **`nohr`** as we did with the **`stcox`** command.

```
. streg gender, dist(exp) nohr          (Can use exp instead of exponential)
```

```
        ( same iterations and output as above )

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
-------------+----------------------------------------------------------------
      gender |    .516186   .0619148      8.337   0.000       .3948352    .6375368
       _cons |    .057755   .0332964      1.735   0.083      -.0075047    .1230147
------------------------------------------------------------------------------
```

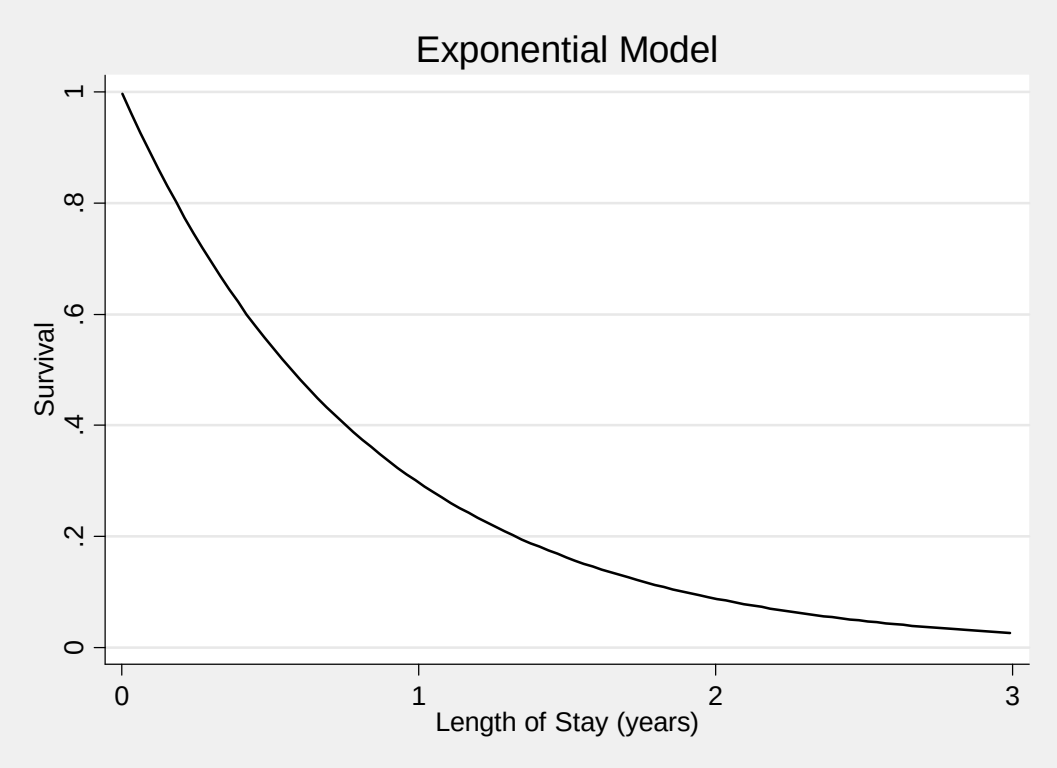
To test whether if there is an association between gender and length of stay we would use the Wald-test:

$$X_W^2 = \frac{(\hat{\beta}_{gender})^2}{\text{var}(\hat{\beta}_{gender})} =$$

(a) Calculate the above Wald test. Is there any association between gender and length of stay?

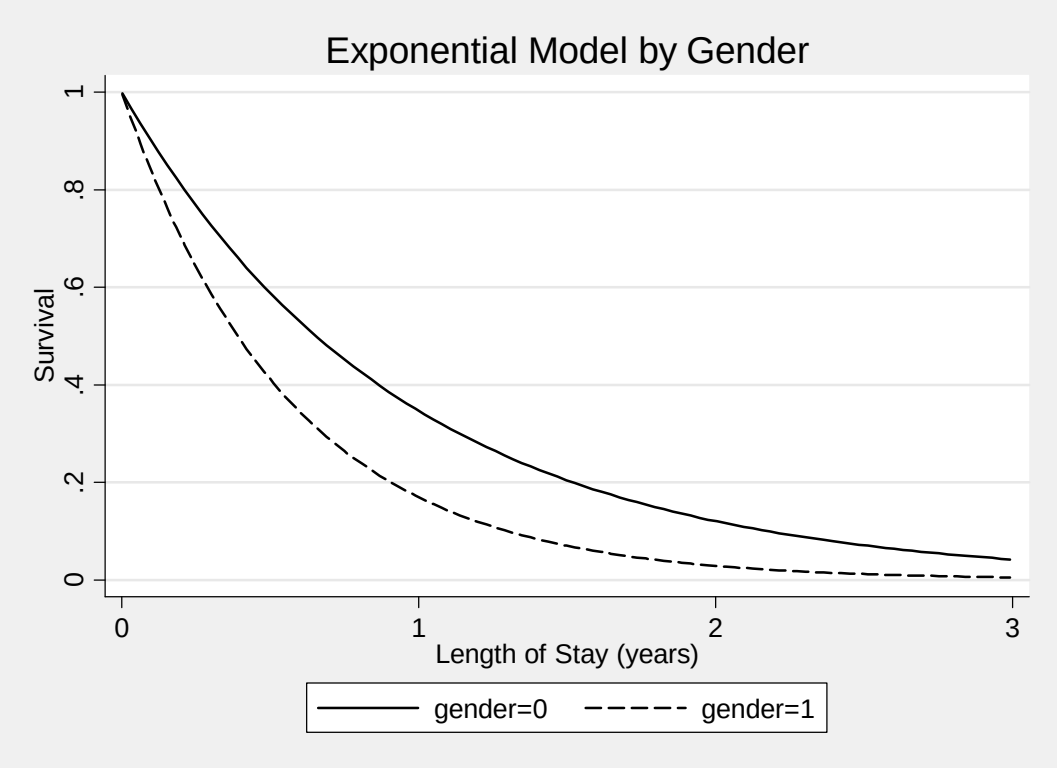
Additionally by using the **`stcurv`** command after **`streg`** we can get a plot of the fitted survival, hazard and cumulative hazard functions by specifying the options **`survival`**, **`hazard`** and **`cumhaz`** each time. So, for example, the survival plot of the exponential model is the following:

```
stcurve, survival title(Exponential Model) xtitle(Length of Stay (years))
```



It is possible to get a separate graph for each group, by giving the option **`at ( )`**. So for females and males we would type the following command:

```
stcurve, survival title(Exponential Model by Gender) xtitle(Length of
Stay (years)) lp(solid dash) lc(black black) at1(gender=0) at2(gender=1)
```



In the same way we can get the Weibull model by specifying the distribution **`weibull:`**

```
. streg gender, dist(weibull)

failure _d:  fail
   analysis time _t:  losyr

Weibull regression -- log relative-hazard form

No. of subjects =         1591                     Number of obs   =      1591
No. of failures =         1269
Time at risk    =  1058.112328
                                                   LR chi2(1)      =     41.73
Log likelihood  =   -3045.2768                     Prob > chi2     =    0.0000

------------------------------------------------------------------------------
      _t | Haz. Ratio   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  gender |   1.512567   .0939336      6.663   0.000       1.339224    1.708347
---------+--------------------------------------------------------------------
   /ln_p |  -.4870456   .0232089    -20.985   0.000      -.5325343   -.4415569
---------+--------------------------------------------------------------------
       p |    .614439   .0142605                          .5871152    .6430345
     1/p |   1.627501   .0377726                          1.555127    1.703243
```

And to get the coefficient estimates:

```
. streg gender, dist(wei) nohr
```

(Again we can use **wei** instead of **weibull**)

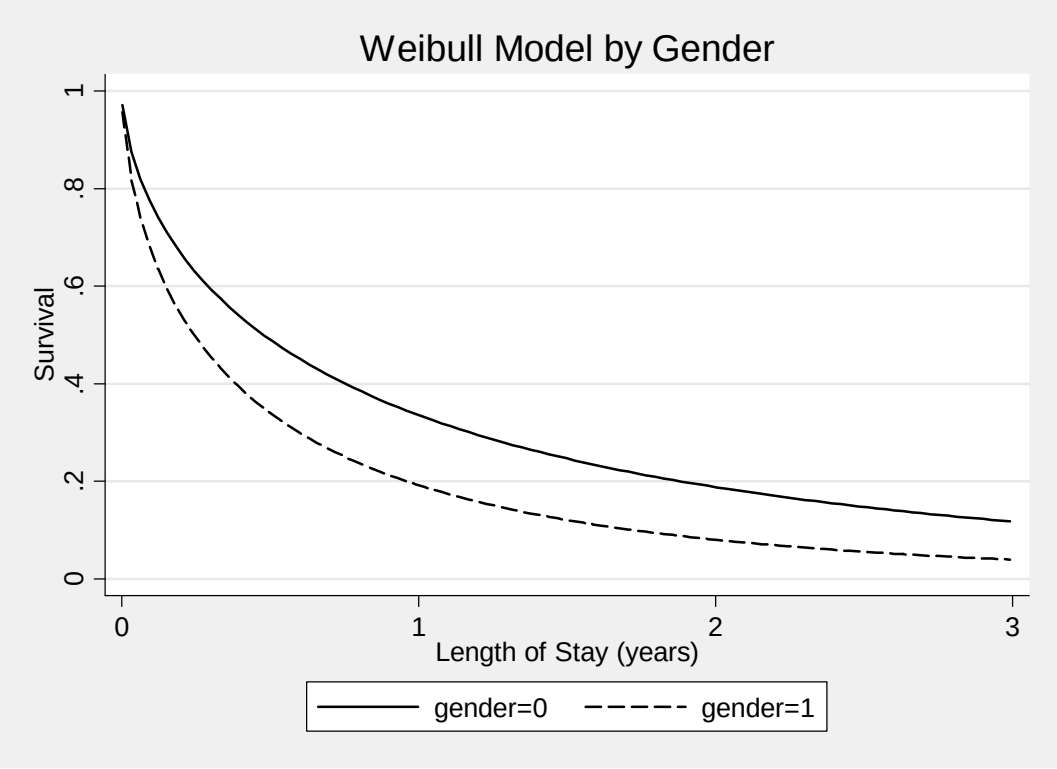
```
No. of subjects =         1591                     Number of obs   =      1591
No. of failures =         1269
Time at risk    =  1058.112328
                                                   LR chi2(1)      =     41.73
Log likelihood  =   -3045.2768                     Prob > chi2     =    0.0000

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
-------------+----------------------------------------------------------------
      gender |   .4138082   .0621021      6.663   0.000       .2920903    .5355261
       _cons |   .0881451   .0333259      2.645   0.008       .0228275    .1534627
-------------+----------------------------------------------------------------
       /ln_p |  -.4870456   .0232089    -20.985   0.000      -.5325343   -.4415569
-------------+----------------------------------------------------------------
           p |    .614439   .0142605                          .5871152    .6430345
         1/p |   1.627501   .0377726                          1.555127    1.703243
------------------------------------------------------------------------------
```

To get the plot of the fitted Weibull model the command is exactly the same as before:

```
. stcurve, survival title(Weibull Model) xtitle(Length of Stay (years))
```

And the graphs by gender:

```
. stcurve, survival title(Weibull Model by Gender) xtitle(Length of
Stay (years)) lp(solid dash) lc(black black) at1(gender=0)
at2(gender=1)
```



(b) From the previous outputs fill in the following table:

Coefficients	***Exponential Model***	***Weibull Model***
β_0		
β_{gender}		
κ		

We know that the survival function of the exponential model is: $S(t) = P(T \geq t) = e^{-\lambda_i t}$ and correspondingly for the Weibull model: $S(t) = P(T \geq t) = e^{-\lambda_i t^k}$, where $\lambda_i = \exp(\beta_0 + \beta_1 gender)$

(c) Calculate the probability of remaining in the nursing home for at least 1 year for each gender based on the fitted exponential and Weibull models.

Then we know that the mean and the median for the exponential model are the following:

$$Mean = 1/\lambda \text{ and } Median = \frac{-\log(0.5)}{\lambda}$$

and for the Weibull distribution the following:

$$Mean = \lambda^{(-1/k)}\Gamma(1/\kappa + 1) \text{ and } Median = \left[\frac{-\log(0.5)}{\lambda}\right]^{1/\kappa}$$

(d) Calculate the mean and median length of stay for each gender according to the exponential and Weibull model. (Note: $\Gamma(2.629) = 1.461$)