13 Introduction to Survival Analysis

In many biomedical studies, the outcome variable is a survival time, or more generally a time to an event. We will describe some of the standard tools for analyzing survival data.

Most studies of survival last a few years, and at completion many subjects may still be alive. For those individuals, the actual survival time is not known – all we know is how long they survived from their entry in the study. Similarly, certain individuals may drop out from the study or be lost to follow-up. Each of these cases is said to be *censored*, and the recorded time for such individuals is their time until the censoring event.

Example: HPA staining for breast cancer survival

We consider data from a retrospective study of 45 women who had surgery for breast cancer. Tumor cells, surgically removed from each woman, were classified according to the results of staining on a marker taken from the Roman snail, the *Helix pomatia agglutinin* (HPA). The marker binds to cancer cells associated with metastasis to nearby lymph nodes. Upon microscopic examination, the cancer cells stained with HPA are classified as positive, corresponding to a tumor with the potential for metastasis, or negative. It is of interest to determine the relationship of HPA staining and the survival of women with breast cancer.

The survival times in months $time_i$ and staining results ($group_i = 0$ for negative and $group_i = 1$ for positive) for the 45 women are presented in the following table. Also included is a *censoring indicator* $cens_i$. Contrary to the normal definition of an indicator variable, the censoring indicator is zero if the observation is right-censored, and one if the observation is uncensored. So it's really a *non-censoring* indicator! A woman's survival time was right censored if the woman was alive at the end of the study or if the woman died of causes unrelated to breast cancer.

	time	group	cens		time	group	cens
1.	23	0	1	24.	40	1	1
2.	47	0	1	25.	41	1	1
3.	69	0	1	26.	48	1	1
4.	70	0	0	27.	50	1	1
5.	71	0	0	28.	59	1	1
6.	100	0	0	29.	61	1	1
7.	101	0	0	30.	68	1	1
8.	148	0	1	31.	71	1	1
9.	181	0	1	32.	76	1	0
10.	198	0	0	33.	105	1	0
11.	208	0	0	34.	107	1	0
12.	212	0	0	35.	109	1	0
13.	224	0	0	36.	113	1	1
14.	5	1	1	37.	116	1	0
15.	8	1	1	38.	118	1	1
16.	10	1	1	39.	143	1	1
17.	13	1	1	40.	154	1	0
18.	18	1	1	41.	162	1	0
19.	24	1	1	42.	188	1	0
20.	26	1	1	43.	212	1	0
21.	26	1	1	44.	217	1	0
22.	31	1	1	45.	225	1	0
23.	35	1	1				

This is the format the data should be in to work with it in **Stata**, but succinctly, the *sorted* survival times for the negative stained women are

$$23, 47, 69, 70^*, 71^*, 100^*, 101^*, 148, 181, 198^*, 208^*, 212^*, 224^*,$$

where $*$ denotes a right-censored observation. The survival times for the positive stained group are

$$5, 8, 10, 13, 18, 24, 26, 26, 31, 35, 40, 41, 48, 50, 59, 61, 68, 71, 76^*, 105^*,$$

$$107^*, 109^*, 113, 116^*, 118, 143, 154^*, 162^*, 188^*, 212^*, 217^*, 225^*.$$

In the breast cancer study, 8 individuals in the negative stained group, and 11 in the positive stained group are censored. Although it is common for studies to have *right-censored* cases, such as we have here, left-censoring and interval-censoring are found in other clinical studies.

Survival Curves

A first step in survival analysis is often to estimate the *survival curve*, or *survival time distribution*. Suppose we are considering a single (homogeneous) population. Let T be the survival time (from some reference point) for a randomly selected individual from the population. Where t is any arbitrary positive value, the survival time distribution is defined to be

$$
\begin{aligned}
S(t) &= Pr(T \geq t) \\
&= \text{probability randomly selected individual survives at least until time } t \\
&= \text{proportion of population that survives at least until time } t.
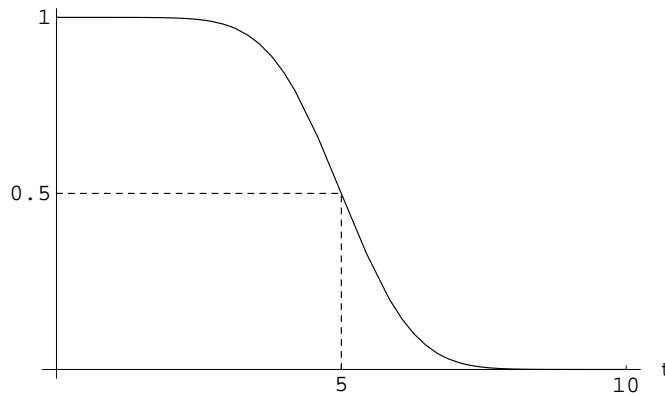\end{aligned}
$$

The function might look like Figure 1.

Figure 1: $S(t)$ versus t;median survival time for population is 5.

Estimating the Survival Curve

Case I: No censoring

If we have a random sample from the population, we use the *empirical survival function*:

$$\hat{S}(t) = \text{sample proportion that survive at least until time t}$$

to estimate $S(t)$. This is easy to compute and plot as a function of t.

Suppose we have a sample of 5 survival times (in days): 5, 8, 20, 30, and 33. $\hat{S}(t)$ has "jumps" of size 1/5 (i.e. 1 divided by the sample size) at each survival time; see Figure 2.

Case II: Right censoring

Recall the data on the survival of women with breast cancer whose cells were negatively stained with HPA:

$$23, 47, 69, 70^*, 71^*, 100^*, 101^*, 148, 181, 198^*, 208^*, 212^*, 224^*,$$

where the $*$ superscript identifies a right-censored observation.

The following algorithm describes the Kaplan-Meier (KM) method for estimating the survival curve (Kaplan-Meier product-limit estimate).
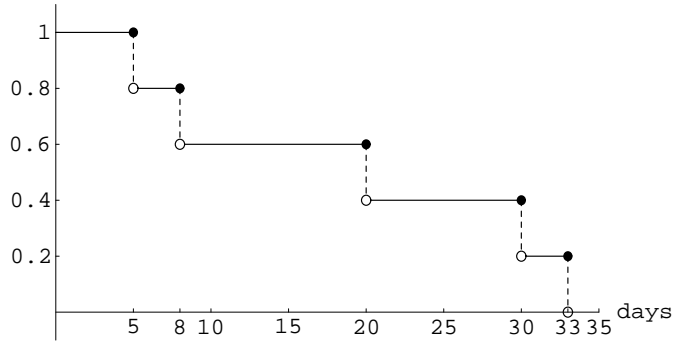
Figure 2: Empirical survival function $\hat{S}(t)$ for the data 5, 8, 20, 30, and 33.

1. Identify times for non-censored cases $0 = t_0 < t_1 < t_2 < \cdots < t_r$. That is, t_1 is the smallest non-censored survival time, t_2 is the second smallest, et cetera. For the example $r = 5$ and $t_0 = 0$, $t_1 = 23$, $t_2 = 47$, $t_3 = 69$, $t_4 = 148$, and $t_5 = 181$.

2. For the jth interval, where $t_{j-1} \leq t < t_j$, evaluate

$$
\begin{aligned}
n_j &= \text{ number at risk (of dying) at beginning of interval,} \\
d_j &= \text{ number of deaths in interval,} \\
\frac{n_j - d_j}{n_j} &= \text{ estimated probability of surviving past } t_{j-1}, \text{ given survival to } t_{j-1} \\
&= \hat{P}(T \geq t_{j-1} | T \geq t_{j-2}).
\end{aligned}
$$

3. For $t_{j-1} \leq t < t_j$,

$$
\begin{aligned}
\hat{S}(t) &= \hat{P}(T \geq t) \\
&= \hat{P}(T \geq t_{j-1} | T \geq t_{j-2}) \times \\
&\quad \hat{P}(T \geq t_{j-2} | T \geq t_{j-3}) \times \cdots \times \\
&\quad \hat{P}(T \geq t_1 | T \geq t_0) \\
&= \frac{n_j - d_j}{n_j} \times \frac{n_{j-1} - d_{j-1}}{n_{j-1}} \times \cdots \times \frac{n_1 - d_1}{n_1}.
\end{aligned}
$$

Remark: Censored observations are taken into account by being treated as cases at risk at the beginning of the interval in which they fail.

To illustrate the calculation for our data, consider the table:

j	Interval	n_j	d_j	$\frac{n_j - d_j}{n_j}$	$\hat{S}(t)$
1	$0 \le t < 23$	13	0	$\frac{13-0}{13} = 1$	1.0
2	$23 \le t < 47$	13	1	$\frac{13-1}{13} = \frac{12}{13} \doteq 0.923$	$1.0 \times 0.923 = 0.923$
3	$47 \le t < 69$	12	1	$\frac{12-1}{12} = \frac{11}{12} \doteq 0.917$	$0.923 \times 0.917 = 0.846$
4	$69 \le t < 148$	11	1	$\frac{10}{11} \doteq 0.909$	$0.846 \times 0.909 = 0.769$
5	$148 \le t < 181$	6	1	$\frac{5}{6} \doteq 0.833$	$0.769 \times 0.833 = 0.641$
6	$181 \le t$	5	1	$\frac{4}{5} = 0.8$	$0.641 \times 0.8 = 0.513$

To obtain the KM estimate in **Stata** we must declare the data we are working with to be survival data. **Stata** then uses the survival time variable and the censoring variable together in analyses. For the breast cancer data we first read in the variables using something like `infile time group cens using c:/breast.txt`. We declare the data to be survival data using `stset time, failure(cens)`. Stata creates several internal variables when we do this. Note that the option `,failure(cens)` makes the variable `cens` into an indicator of known death ("failure"). Finally we obtain the KM survival curve estimates across the two groups with the command `sts graph, by(group)`. In Figure 3 we have a picture of $\hat{S}(t)$ from the negatively stained group as well as the estimate from the positively stained group. Note that the negatively stained group tends to live longer, as we would expect.
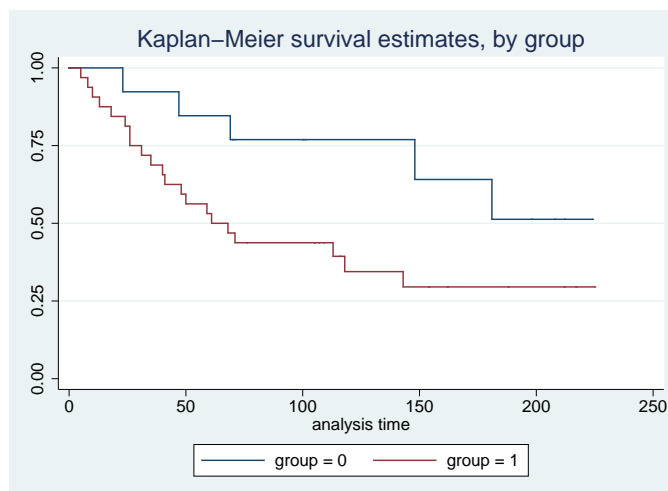

Figure 3: KM survival curves for positively and negatively stained groups.

The estimated quartiles for survival across the two groups are obtained by `stsum, by(group)`. Annotated output follows; for example, we see that the median survival in the positive stained group is estimated to be 61 months.

```
. stsum,by(group)
         failure _d:  cens
   analysis time _t:  time
             |                  incidence      no. of   |------ Survival time -----|
group        | time at risk       rate       subjects        25%       50%       75%
---------+----------------------------------------------------------------------------
       0 |         1652    .0030266             13          148         .         .
       1 |         2679    .0078387             32           26        61         .
---------+----------------------------------------------------------------------------
   total |         4331    .0060032             45           40       113         .
```

Some remarks:

- The estimated survival curve "drops to zero" only if the last case is not censored.

- The KM curve allows us to estimate percentiles of the survival distribution, with a primary interest being the median survival time (50th percentile). In the example above (group 0), the 90th percentile is approximately 47 months (i.e. we estimate that 90% of the population will survive at least 47 months). The median cannot be estimated here – all we can say is that we estimate the median to be at least 181 months.

- The KM estimate is the usual empirical estimate if no cases are censored.

- Statistical methods are available to

  - Estimate the mean survival time. This sounds good, but survival time distributions tend to be highly skewed right, so usually we are much more interested in the median.

    ```
    . stci,by(group) rmean
             failure _d:  cens
       analysis time _t:  time
                 |    no. of   restricted
    group        |  subjects        mean      Std. Err.     [95% Conf. Interval]
    -----------+------------------------------------------------------------------
           0 |        13    167.7436(*)    20.80779       126.961       208.526
           1 |        32    104.0477(*)     15.6278       73.4177       134.678
    -----------+------------------------------------------------------------------
       total |        45    121.9075(*)     13.3793       95.6846       148.13
    (*) largest observed analysis time is censored, mean is underestimated.

    . stci,by(group) emean
             failure _d:  cens
       analysis time _t:  time
                 |    no. of     extended
    group        |  subjects         mean
    -----------+----------------------------
           0 |        13     339.7513
           1 |        32     158.5235
    -----------+----------------------------
       total |        45     196.3361
    ```
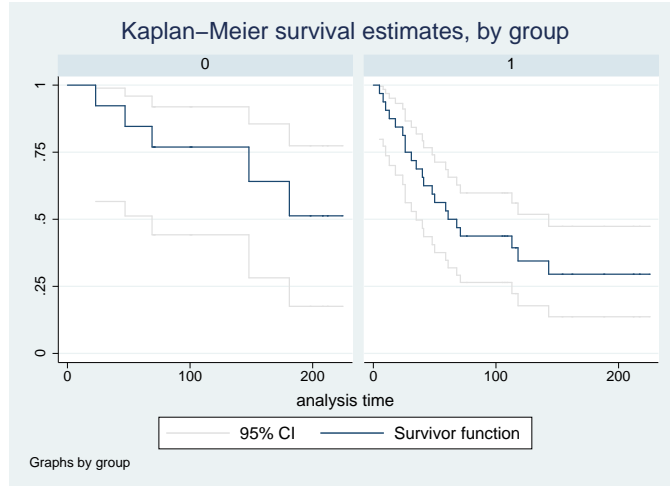
  - Get a C.I. for the survival curve. You need to ask for pointwise Greenwood confidence bands.

    ```
    . sts graph,by(group) gwood
             failure _d:  cens
       analysis time _t:  time
    ```

Kaplan–Meier survival estimates, by group

– Compare survival curves across groups – you can think of this as the censored data analogue of (non-parametric) ANOVA. There are a lot of available tests, though most common probably is the log-rank test (Stata's default). A few of them follow.

```
. sts test group
          failure _d:  cens
    analysis time _t:  time

Log-rank test for equality of survivor functions
        |    Events         Events
group   |  observed       expected
--------+--------------------------
0       |         5           9.57
1       |        21          16.43
--------+--------------------------
Total   |        26          26.00
            chi2(1) =       3.51
            Pr>chi2 =     0.0608
. sts test group,wilc
          failure _d:  cens
    analysis time _t:  time

Wilcoxon (Breslow) test for equality of survivor functions
        |    Events         Events        Sum of
group   |  observed       expected         ranks
--------+----------------------------------------
0       |         5           9.57          -159
1       |        21          16.43           159
--------+----------------------------------------
Total   |        26          26.00             0
            chi2(1) =       4.18
            Pr>chi2 =     0.0409
. sts test group,cox
          failure _d:  cens
    analysis time _t:  time

Cox regression-based test for equality of survival curves
        |    Events         Events      Relative
group   |  observed       expected        hazard
--------+----------------------------------------
0       |         5           9.57        0.5633
1       |        21          16.43        1.3966
--------+----------------------------------------
Total   |        26          26.00        1.0000
         LR chi2(1) =       3.87
            Pr>chi2 =     0.0491
. sts test group,peto
```

```
              failure _d:  cens
        analysis time _t:  time
     Peto-Peto test for equality of survivor functions
            |    Events        Events        Sum of
     group  |  observed      expected         ranks
     -------+------------------------------------------
     0      |         5          9.57     -3.509069
     1      |        21         16.43      3.509069
     -------+------------------------------------------
     Total  |        26         26.00             0
                    chi2(1) =       4.12
                    Pr>chi2 =     0.0425
     . sts test group,tware
              failure _d:  cens
        analysis time _t:  time
     Tarone-Ware test for equality of survivor functions
            |    Events        Events        Sum of
     group  |  observed      expected         ranks
     -------+------------------------------------------
     0      |         5          9.57    -26.921999
     1      |        21         16.43     26.921999
     -------+------------------------------------------
     Total  |        26         26.00             0
                    chi2(1) =       4.05
                    Pr>chi2 =     0.0441
```

These tests do not necessarily all agree with each other – some emphasize different parts of the distribution than others, and different principles for approximating distributions are employed. In this case there is little difference, except the minor differences are right at the 0.05 significance level.

The Cox Proportional Hazards Model

The risk of failing at time t is defined to be the probability of an individual dying in the "next instant" (e.g. in a time frame of length Δ) given this individual has survived at least until time t:

$$P(t \leq T < t + \Delta | t \leq T).$$

We define the *hazard function* $h(t)$ such that for small enough Δ,

$$P(t \leq T < t + \Delta | t \leq T) = h(t)\Delta.$$

The hazard function is proportional to the instantaneous "risk of failing" at any time t, given that an individual has lived at least to time t.

Now consider two individuals, 1 and 2, each with their own hazard functions $h_1(t)$ and $h_2(t)$. If we assume that one individual's instantaneous rate of failing is a constant multiple of the other's, i.e. $h_2(t) = ah_1(t)$ for some constant a, then these two individuals have *proportional hazard functions*. Figure 4 shows an example of this phenomenon where the hazard ratio is $1/2$.

Proportional hazards may or may not be a reasonable assumption to make. For example, consider two people, roughly the same age and demographic except that at the age of 20, person 2 takes up smoking while person 1 does not. You will hopefully agree with me that initially, the smoker and the non-smoker will most likely have *identical* hazards. As the years roll by, and smoking takes its toll, we would think that the smoker's instantaneous rate of failing, which is proportional to the probability of dying in the next minute, say, will increase relative to the hazard for the non-smoker. In this example proportional hazards probably is an unreasonable assumption.

The proportional hazards *model* generalizes the above concept for n individuals, each with their own covariate value x_i or set of p covariate values $x_i = (x_{i1}, x_{i2}, \ldots, x_{ip})$. In the case where the
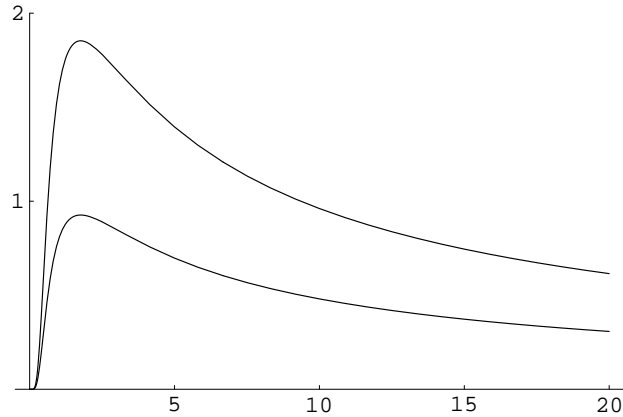
Figure 4: An example of proportional hazard functions; here the constant of proportionality is 0.5.

n individuals only have one covariate, the model stipulates for individuals i and j, with hazard functions $h_i(t)$ and $h_j(t)$ respectively, that

$$h_i(t)e^{-\beta x_i} = h_j(t)e^{-\beta x_j}.$$

Note that this implies

$$\frac{h_i(t)}{h_j(t)} = \frac{e^{\beta x_i}}{e^{\beta x_j}} = e^{\beta(x_i - x_j)}.$$

Here, $e^{\beta(x_i - x_j)}$ is the relative risk of instantaneous failure at *any time t* for individuals i and j. That is the power of the proportional hazards assumption: the relative risk of dying for two individuals is a simple function of the model parameters and holds for all t, independent of the value of t. If individual i has covariate value $x + 1$ and individual j has covariate value x, i.e. their covariate values only differ by 1 unit on the covariate measurement scale, then

$$\frac{h_i(t)}{h_j(t)} = \frac{e^{\beta(x+1)}}{e^{\beta x}} = e^{\beta}.$$

Thus, e^{β} is the relative risk of failing in the next instant when we increase the covariate by one unit. Note that if x_i is a simple zero/one variable denoting which group individual i falls into, then e^{β} is the relative risk of failing in the next instant for the group denoted by $x_i = 1$ versus $x_i = 0$.

The Cox PH model is fit as follows (everything had to be `stset` prior to this command). We do not have to specify the dependent variable as in other regression routines because it is defined (survival time) with `stset`.

```
. stcox group
        failure _d:  cens
   analysis time _t:  time
Iteration 0:   log likelihood = -86.983777
Iteration 1:   log likelihood = -85.087844
Iteration 2:   log likelihood = -85.048003
Iteration 3:   log likelihood = -85.047944
Refining estimates:
Iteration 0:   log likelihood = -85.047944
Cox regression -- Breslow method for ties
No. of subjects =           45                    Number of obs   =          45
No. of failures =           26
Time at risk    =         4331
```

```
                                              LR chi2(1)      =       3.87
Log likelihood  =    -85.047944               Prob > chi2     =     0.0491
------------------------------------------------------------------------------
        _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----------+------------------------------------------------------------------
     group |   2.479398   1.241987     1.81   0.070     .9288808    6.618086
------------------------------------------------------------------------------
**************
************* Note:  stcox group,nohr  reports coefficients, not hazard ratios
------------------------------------------------------------------------------
        _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----------+------------------------------------------------------------------
     group |   .9080157   .5009228     1.81   0.070    -.0737749    1.889806
------------------------------------------------------------------------------
```

We have an estimate of $\hat{\beta} = 0.908$ and the estimated relative risk is $e^{\hat{\beta}} = e^{0.908} \doteq 2.5$. That is, those with positive staining are estimated to have a risk of dying in the next instant about 2.5 times as great as those with negative staining. Note that the p-value for $H_0 : \beta = 0$ is small but not significant at the 5% level. There is definitely *some* indication that staining affects survival, with positive staining decreasing survival. A 95% C.I. for the risk may be obtained by exponentiating the endpoints for the C.I. for β. Here, we estimate the relative risk of expiring (for positive compared to negative staining) to be within $(e^{-0.073}, e^{1.89}) = (0.93, 6.62)$ with 95% confidence.

Remark: The hazard function for individual i can be defined to be a scale multiple $e^{x_i\beta}$ of a *baseline* hazard function denoted $h_0(t)$. The model may be recast as $h_i(t) = h(t|x_i) = e^{x_i\beta}h_0(t)$. This baseline hazard function $h_0(t)$ and β together thus completely determine the model. The baseline hazard $h_0(t)$ may be estimated from the data as well as survival curves, median and mean survival, et cetera, for any covariate value x. These sorts of inferences are quite easy to get out of Stata but a bit beyond what is comfortable to cover in this class.

Checking the Proportional Hazards Assumption

The assumption matters because estimates and inferences based on them may be completely invalid if hazards are not proportional, and we have no easy way to assess that from the regression output. There are standard diagnostics that often work pretty well but are not at all easy to derive without some mathematics.

What you would like to do is estimate the hazard functions from the data, and plot them on the same scale to assess proportionality. That is very similar to estimating a population frequency distribution, and is a fairly hard problem that takes quite a bit of data (Stata will do it, but that doesn't mean it works well). What works better and is much easier to estimate (effectively from Kaplan-Meier) is the integrated or cumulative hazard $H(t) = \int_0^t h(s)ds$. If hazard functions are proportional then so are cumulative hazard functions. A very fortunate mathematical relationship is that the survival curve is related to the cumulative hazard as $H(t) = -\log S(t)$ so $\log H(t) = \log\left[-\log S(t)\right]$. If $h_0(t)$ is the baseline hazard function, then the proportional hazards assumptions says the cumulative hazard function for individual i is $e^{x_i\beta}H_0(t)$, and $\log\left[-logS_i(t)\right] = x_i\beta + \log\left[-logS_0(t)\right]$. What this says is that if x is a group indicator as above then plots of $\log\left[-\log S(t)\right]$ for groups should be parallel (and separated by the β's for groups). Stata does (almost) this in its `stphplot` command.

Another approach is to compare the completely nonparametric Kaplan-Meier curve to the curve predicted under proportional hazards. That is the `stcoxkm` command. Stata's description of these commands from the help system is as follows:

Description

> stphplot plots -ln{-ln(survival)} curves for each category of a nominal or ordinal covariate versus ln(analysis time). These are often referred to as "log-log" plots. Optionally, these estimates can be adjusted for covariates. The proportional-hazards assumption is not violated when the curves are parallel.

> stcoxkm plots Kaplan-Meier observed survival curves and compares them to the Cox predicted curves for the same variable. The closer the observed values are to the predicted, the less likely it is that the proportional-hazards assumption has been violated. Do not run stcox prior to this command; stcoxkm will execute stcox itself to estimate the model and obtain predicted values.

Doing this with the breast cancer data using these commands yields the following graph:

```
. stphplot,by(group) name(loglog)
. stcoxkm,by(group) name(phkm)
. graph combine loglog phkm
```
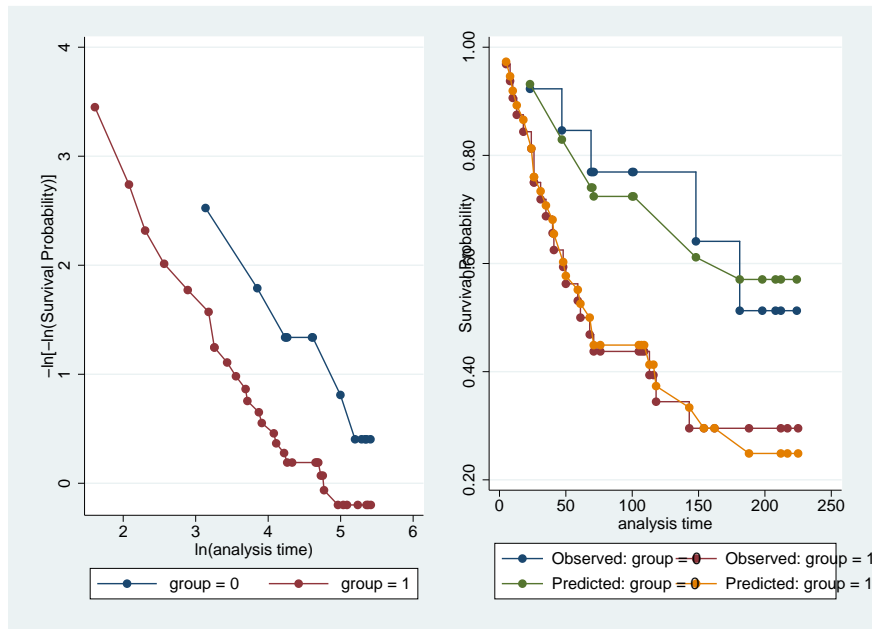

This all looks great. We do not have any evidence that the proportional hazards assumption is not reasonable.

A final example

We examine a data set consisting of the time spent running on a treadmill for 14 people aged 15 and older. Each subject's gender and age were recorded. It is of interest to the experimenter how age and gender affect ones endurance.

When fitting the PH model with gender and age as main effects,

$$h(t|\text{age, gender}) = e^{age \times \beta_1 + gender \times \beta_2} h_0(t),$$

we are going to let `xi` determine the gender indicator (it will set the first group alphabetically, females, to 0 and males to 1). The baseline group (i.e. those with covariates age = 0 and gender indicator = 0, and thus a hazard function of $e^{0\beta_1 + 0\beta_2} = e^0 h_0(t) = h_0(t)$) consists of females of age zero, which is not interpretable in this context. Observations were censored due to a subject having to leave the treadmill for reasons other than being tired. The data follow:

```
         gender    age    minutes    cens    weight
  1.       male     34         16       1       215
  2.       male     15         35       0       135
  3.     female     22         55       0       145
  4.     female     18         95       1        97
  5.       male     18         55       0       225
  6.     female     32         55       1       185
  7.     female     37         25       1       155
  8.     female     67         15       1       142
  9.     female     55         22       1       132
 10.       male     55         13       1       183
 11.       male     62         13       1       168
 12.     female     33         57       0       132
 13.     female     17         52       0       112
 14.       male     24         54       1       175
```

We need to start by declaring a survival data set:

```
. stset minutes,failure(cens)
     failure event:  cens != 0 & cens < .
obs. time interval:  (0, minutes]
 exit on or before:  failure
-------------------------------------------------------------------------------
       14  total obs.
        0  exclusions
-------------------------------------------------------------------------------
       14  obs. remaining, representing
        9  failures in single record/single failure data
      562  total analysis time at risk, at risk from t =          0
                                earliest observed entry t =          0
                                 last observed exit t =         95
```

The fit of the model with only gender $h(t|\text{gender}) = e^{\text{gender indicator} \times \beta_1} h_0(t)$:

```
. xi: stcox i.gender
i.gender          _Igender_1-2          (_Igender_1 for gender==female omitted)
        failure _d:  cens
   analysis time _t:  minutes
Iteration 0:   log likelihood = -18.061924
Iteration 1:   log likelihood = -17.622711
Iteration 2:   log likelihood = -17.620368
Refining estimates:
Iteration 0:   log likelihood = -17.620368
Cox regression -- Breslow method for ties
No. of subjects =           14                   Number of obs    =          14
No. of failures =            9
Time at risk    =          562
                                                 LR chi2(1)       =        0.88
Log likelihood  =   -17.620368                   Prob > chi2      =      0.3474
-------------------------------------------------------------------------------
        _t | Haz. Ratio   Std. Err.       z    P>|z|     [95% Conf. Interval]
-----------+-------------------------------------------------------------------
 _Igender_2 |   1.971276    1.411726    0.95   0.343     .4843518    8.022949
-------------------------------------------------------------------------------
```

The test for a gender effect yields a p-value of 0.343. We would accept at any reasonable significance level that there is not a gender effect. The estimated hazard ratio is

$$h(t|\text{gender} = \text{male})/h(t|\text{gender} = \text{female}) = 1.97$$

for all t. That is, the probability of a randomly picked man failing (stepping off the treadmill) in the next second is estimated be twice the probability of a randomly picked female. The confidence interval for the hazard ratio is from 0.48 to 8.02, which includes 1 since the effect was not significant.

Let's look at the model fit with only age $h(t|\text{age}) = e^{\text{age} \times \beta_1} h_0(t)$:

```
. stcox age
        failure _d:  cens
```

```
     analysis time _t:  minutes
Iteration 0:   log likelihood = -18.061924
Iteration 1:   log likelihood = -11.184791
Iteration 2:   log likelihood = -11.184559
Refining estimates:
Iteration 0:   log likelihood = -11.184559
Cox regression -- Breslow method for ties
No. of subjects =            14                 Number of obs   =          14
No. of failures =             9
Time at risk    =           562
                                                LR chi2(1)      =       13.75
Log likelihood  =   -11.184559                  Prob > chi2     =      0.0002
-------------------------------------------------------------------------------
        _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----------+-------------------------------------------------------------------
       age |   1.118133    .043125     2.90   0.004     1.036725    1.205934
-------------------------------------------------------------------------------
```

A year from now, a randomly selected individual will be 1.118 times as likely to step off the treadmill after 15 minutes (or any amount of time) than now. In ten years it will be $1.118^{10} = 3.05$ times as likely. When we fit the model with both of these predictors $h(t|\text{age, gender}) = e^{\text{age} \times \beta_1 + \text{gender indicator} \times \beta_2} h_0(t)$ we see that estimated regression effects, and therefore model interpretation, change somewhat:

```
. xi: stcox i.gender age
i.gender          _Igender_1-2       (_Igender_1 for gender==female omitted)
      failure _d:  cens
   analysis time _t:  minutes
Iteration 0:   log likelihood = -18.061924
Iteration 1:   log likelihood = -8.3270231
Iteration 2:   log likelihood = -7.2765366
Iteration 3:   log likelihood = -7.1238759
Iteration 4:   log likelihood = -7.1166297
Iteration 5:   log likelihood = -7.1166049
Refining estimates:
Iteration 0:   log likelihood = -7.1166049
Cox regression -- Breslow method for ties
No. of subjects =            14                 Number of obs   =          14
No. of failures =             9
Time at risk    =           562
                                                LR chi2(2)      =       21.89
Log likelihood  =   -7.1166049                  Prob > chi2     =      0.0000
-------------------------------------------------------------------------------
        _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----------+-------------------------------------------------------------------
 _Igender_2 |   34.87809   55.05714     2.25   0.024     1.580812    769.529
       age |   1.244367   .1064681     2.56   0.011     1.052251    1.471558
-------------------------------------------------------------------------------
```

At a given age, a random male running alongside a random female is about 35 times as likely to step off the treadmill at any time. A woman 20 years older than another woman is about $1.244^{20} \doteq 79$ times as likely to step off compared to the younger woman. Note that in the presence of age, gender is now significant, although by itself gender is not a significant factor. In this case age is said to be a *suppressor* variable.
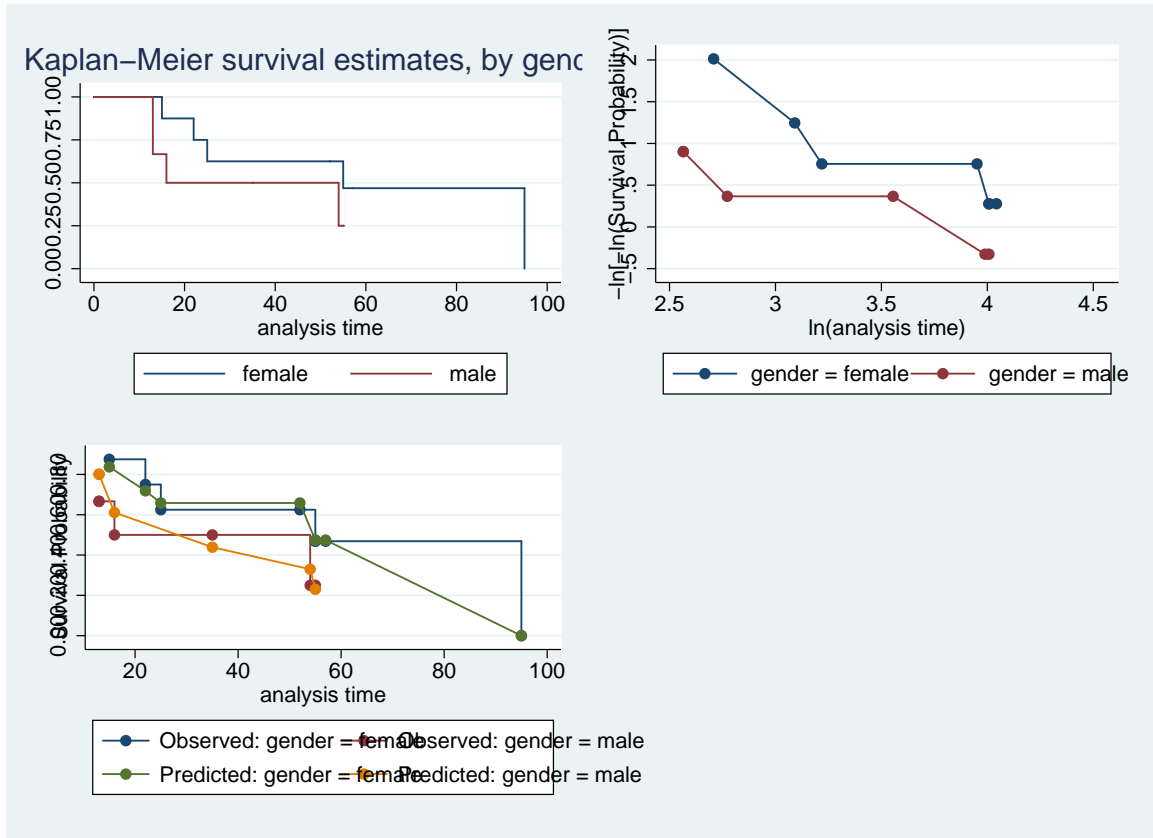
In the model fit that included an interaction between age and gender, the interaction term was not significant. Weight is not significant in the presence of gender and age (or by itself).

We should check on the proportional hazards assumption.

```
. sts graph,by(gender) name(km)
. stphplot,by(gender) name(loglog)
. stcoxkm,by(gender) name(phkm)
. graph combine km loglog phkm
```

There is not a lot of data here, so we have to expect deviations from ideal patterns. The log-log plot only needs to be parallel, not linear, and this is plausibly parallel. The PH model is not disagreeing with the Kaplan-Meier fit. The survival curves by gender are consistent with the tests we ran. There is no clear reason to question the PH assumptions.