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 \( t_i \) and staining results (group \( g_i = 0 \) for negative and group \( g_i = 1 \) 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.

<table>
<thead>
<tr>
<th>time</th>
<th>group</th>
<th>cens</th>
<th>time</th>
<th>group</th>
<th>cens</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>23</td>
<td>0</td>
<td>24.</td>
<td>40</td>
<td>1</td>
</tr>
<tr>
<td>3.</td>
<td>47</td>
<td>0</td>
<td>25.</td>
<td>41</td>
<td>1</td>
</tr>
<tr>
<td>4.</td>
<td>69</td>
<td>0</td>
<td>26.</td>
<td>48</td>
<td>1</td>
</tr>
<tr>
<td>5.</td>
<td>70</td>
<td>0</td>
<td>27.</td>
<td>50</td>
<td>1</td>
</tr>
<tr>
<td>6.</td>
<td>71</td>
<td>0</td>
<td>28.</td>
<td>59</td>
<td>1</td>
</tr>
<tr>
<td>7.</td>
<td>100</td>
<td>0</td>
<td>29.</td>
<td>61</td>
<td>1</td>
</tr>
<tr>
<td>8.</td>
<td>101</td>
<td>0</td>
<td>30.</td>
<td>68</td>
<td>1</td>
</tr>
<tr>
<td>9.</td>
<td>181</td>
<td>0</td>
<td>31.</td>
<td>71</td>
<td>1</td>
</tr>
<tr>
<td>10.</td>
<td>198</td>
<td>0</td>
<td>32.</td>
<td>76</td>
<td>0</td>
</tr>
<tr>
<td>11.</td>
<td>208</td>
<td>0</td>
<td>33.</td>
<td>105</td>
<td>0</td>
</tr>
<tr>
<td>12.</td>
<td>212</td>
<td>0</td>
<td>34.</td>
<td>107</td>
<td>1</td>
</tr>
<tr>
<td>13.</td>
<td>224</td>
<td>0</td>
<td>35.</td>
<td>109</td>
<td>0</td>
</tr>
<tr>
<td>14.</td>
<td>225</td>
<td>1</td>
<td>36.</td>
<td>113</td>
<td>1</td>
</tr>
<tr>
<td>15.</td>
<td>226</td>
<td>1</td>
<td>37.</td>
<td>116</td>
<td>0</td>
</tr>
<tr>
<td>16.</td>
<td>231</td>
<td>1</td>
<td>38.</td>
<td>118</td>
<td>1</td>
</tr>
<tr>
<td>17.</td>
<td>235</td>
<td>1</td>
<td>39.</td>
<td>143</td>
<td>1</td>
</tr>
<tr>
<td>18.</td>
<td>236</td>
<td>1</td>
<td>40.</td>
<td>154</td>
<td>0</td>
</tr>
<tr>
<td>19.</td>
<td>243</td>
<td>1</td>
<td>41.</td>
<td>162</td>
<td>0</td>
</tr>
<tr>
<td>20.</td>
<td>246</td>
<td>1</td>
<td>42.</td>
<td>212</td>
<td>0</td>
</tr>
<tr>
<td>21.</td>
<td>268</td>
<td>1</td>
<td>43.</td>
<td>217</td>
<td>0</td>
</tr>
<tr>
<td>22.</td>
<td>275</td>
<td>1</td>
<td>44.</td>
<td>225</td>
<td>1</td>
</tr>
<tr>
<td>23.</td>
<td>321</td>
<td>1</td>
<td>45.</td>
<td>225</td>
<td>1</td>
</tr>
</tbody>
</table>

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

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

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

\[
S(t) = \Pr(T \geq t)
\]

= probability randomly selected individual survives at least until time \( t \)

= proportion of population that survives at least until time \( t \).

The function might look like Figure 1.

![Figure 1: \( S(t) \) versus \( t \); median survival time for population is 5.](image)

**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:

\[
\]

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).
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 $j^{th}$ interval, where $t_{j-1} \leq t < t_j$, evaluate

\[
\frac{n_j - d_j}{n_j} \quad \text{estimated probability of surviving past } t_{j-1}, \text{ given survival to } t_{j-1} = \hat{P}(T \geq t_{j-1} | T \geq t_{j-2}).
\]

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

\[
\hat{S}(t) = \hat{P}(T \geq t) = \hat{P}(T \geq t_{j-1}|T \geq t_{j-2}) \times \hat{P}(T \geq t_{j-2}|T \geq t_{j-3}) \times \cdots \times \hat{P}(T \geq t_1|T \geq t_0) = \frac{n_j - d_j}{n_j} \times \frac{n_{j-1} - d_{j-1}}{n_{j-1}} \times \cdots \times \frac{n_1 - d_1}{n_1}.
\]

**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:
13 INTRODUCTION TO SURVIVAL ANALYSIS

<table>
<thead>
<tr>
<th>$j$</th>
<th>Interval</th>
<th>$n_j$</th>
<th>$d_j$</th>
<th>$\frac{n_j - d_j}{n_j}$</th>
<th>$\hat{S}(t)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$0 \leq t &lt; 23$</td>
<td>13</td>
<td>0</td>
<td>$\frac{13-0}{13} = 1$</td>
<td>1.0</td>
</tr>
<tr>
<td>2</td>
<td>$23 \leq t &lt; 47$</td>
<td>13</td>
<td>1</td>
<td>$\frac{13-1}{13} = \frac{12}{13} \approx 0.923$</td>
<td>$1.0 \times 0.923 = 0.923$</td>
</tr>
<tr>
<td>3</td>
<td>$47 \leq t &lt; 69$</td>
<td>12</td>
<td>1</td>
<td>$\frac{12-1}{12} = \frac{11}{12} \approx 0.917$</td>
<td>$0.923 \times 0.917 = 0.846$</td>
</tr>
<tr>
<td>4</td>
<td>$69 \leq t &lt; 148$</td>
<td>11</td>
<td>1</td>
<td>$\frac{10}{11} \approx 0.909$</td>
<td>$0.846 \times 0.909 = 0.769$</td>
</tr>
<tr>
<td>5</td>
<td>$148 \leq t &lt; 181$</td>
<td>6</td>
<td>1</td>
<td>$\frac{5}{6} \approx 0.833$</td>
<td>$0.769 \times 0.833 = 0.641$</td>
</tr>
<tr>
<td>6</td>
<td>$181 \leq t$</td>
<td>5</td>
<td>1</td>
<td>$\frac{4}{5} = 0.8$</td>
<td>$0.641 \times 0.8 = 0.513$</td>
</tr>
</tbody>
</table>

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 `in file time group cens using c:/breast.txt`. We declare the data to be survival data using `st set 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.

![Kaplan-Meier survival estimates, by group](image)

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

The estimated quartiles for survival across the two groups are obtained by `st sum, 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: censanalysis 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: censanalysis time _t: time  

| no. of restricted  | mean Std. Err. [95% Conf. Interval]  
| 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: censanalysis time _t: time  

| no. of extended  | mean  
| 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: censanalysis time _t: time
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

<table>
<thead>
<tr>
<th>Events</th>
<th>Events</th>
</tr>
</thead>
<tbody>
<tr>
<td>group</td>
<td>observed</td>
</tr>
<tr>
<td>--------</td>
<td>-----------</td>
</tr>
<tr>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>21</td>
</tr>
<tr>
<td>Total</td>
<td>26</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Events</th>
<th>Events</th>
<th>Sum of</th>
</tr>
</thead>
<tbody>
<tr>
<td>group</td>
<td>observed</td>
<td>expected</td>
</tr>
<tr>
<td>--------</td>
<td>-----------</td>
<td>-----------</td>
</tr>
<tr>
<td>0</td>
<td>5</td>
<td>9.57</td>
</tr>
<tr>
<td>1</td>
<td>21</td>
<td>16.43</td>
</tr>
<tr>
<td>Total</td>
<td>26</td>
<td>26.00</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Events</th>
<th>Events</th>
<th>Relative</th>
</tr>
</thead>
<tbody>
<tr>
<td>group</td>
<td>observed</td>
<td>expected</td>
</tr>
<tr>
<td>--------</td>
<td>-----------</td>
<td>-----------</td>
</tr>
<tr>
<td>0</td>
<td>5</td>
<td>9.57</td>
</tr>
<tr>
<td>1</td>
<td>21</td>
<td>16.43</td>
</tr>
<tr>
<td>Total</td>
<td>26</td>
<td>26.00</td>
</tr>
</tbody>
</table>

LR chi2(1) = 3.87
Pr>chi2 = 0.0491
```

Graphs by group

Kaplan–Meier survival estimates, by group
13 Introduction to Survival Analysis

failure _d: cens
analysis time _t: time

Peto-Peto test for equality of survivor functions

<table>
<thead>
<tr>
<th>group</th>
<th>Events observed</th>
<th>Events expected</th>
<th>Sum of ranks</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5</td>
<td>9.57</td>
<td>-3.509069</td>
</tr>
<tr>
<td>1</td>
<td>21</td>
<td>16.43</td>
<td>3.509069</td>
</tr>
<tr>
<td>Total</td>
<td>26</td>
<td>26.00</td>
<td>0</td>
</tr>
</tbody>
</table>

\[ \chi^2(1) = 4.12 \]
\[ \text{Pr}>\chi^2 = 0.0425 \]

. ststest group, tware

failure _d: cens
analysis time _t: time

Tarone-Ware test for equality of survivor functions

<table>
<thead>
<tr>
<th>group</th>
<th>Events observed</th>
<th>Events expected</th>
<th>Sum of ranks</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5</td>
<td>9.57</td>
<td>-26.921999</td>
</tr>
<tr>
<td>1</td>
<td>21</td>
<td>16.43</td>
<td>26.921999</td>
</tr>
<tr>
<td>Total</td>
<td>26</td>
<td>26.00</td>
<td>0</td>
</tr>
</tbody>
</table>

\[ \chi^2(1) = 4.05 \]
\[ \text{Pr}>\chi^2 = 0.0441 \]

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 \( \Delta \)) 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 \( \Delta \),

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

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, $e^{\beta}$ 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 $e^{\beta}$ 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
```

13  INTRODUCTION TO SURVIVAL ANALYSIS
We have an estimate of $\hat{\beta} = 0.908$ and the estimated relative risk is $e^{\hat{\beta}} = e^{0.908} \approx 2.5$. 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 $\beta$. 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 $\beta$ 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 $\log H(t) = \log [-\log S(t)]$. 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 $\log [-\log S_i(t)] = x_i \beta + \log [-\log S_0(t)]$. What this says is that if $x$ is a group indicator as above then plots of $\log [-\log S(t)]$ for groups should be parallel (and separated by the $\beta$’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 stcoxxkm command. Stata’s description of these commands from the help system is as follows:
Description

\texttt{stphplot} plots \(-\ln\{-\ln(\text{survival})\}\) curves for each category of a nominal or ordinal covariate versus \(\ln(\text{analysis time})\). 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.

\texttt{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 \texttt{stcox} prior to this command; \texttt{stcoxkm} will execute \texttt{stcox} itself to estimate the model and obtain predicted values.

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

\begin{verbatim}
. stphplot, by(group) name(loglog)
. stcoxkm, by(group) name(phkm)
. graph combine loglog phkm
\end{verbatim}

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,

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

we are going to let \(x_i\) 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:
We need to start by declaring a survival data set:

```
. stset minutes, failure(cens)
```

The fit of the model with only gender

\[ h(t|\text{gender}) = e^{\text{gender indicator} \times \beta_1} h_0(t) \]

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

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

\[ h(t|\text{age}) = e^{\text{age} \times \beta_1} h_0(t) \]
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 $h(t|\text{age}, \text{gender}) = e^{\text{age} \times \beta_1 + \text{gender indicator} \times \beta_2}h_0(t)$ we see that estimated regression effects, and therefore model interpretation, change somewhat:

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 $1.244^{20} = 79$ 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).
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.