23.August 2010

Chapter 4

Regression Models

- The target variable T denotes failure time

- We let $\underline{x} = (x^{(1)}, \ldots, x^{(m)})'$ represent a vector of available **covariates**.

  Also called regression variables, regressors, prognostic factors, or explanatory variables.

- We are interested in modelling and determining the relationship between T and $\underline{x}$.

  Often this is referred to as **prognostic factor analysis**.

- **The primary question is:** Do any subsets of the m covariates help to explain survival time?

  For example, does age at first treatment and/or gender increase or decrease (relative) risk of survival? If so, how and by what estimated quantity?

  Does one treatment give a significant survival advantage over the other treatment?

Example 1. Let

- $x^{(1)}$ denotes the sex $(x_i^{(1)} = 1$ for males and
$$x_i^{(1)} = 0 \text{ for females}),$$

- $x^{(2)} = $ Age at diagnosis,

- $x^{(3)} = x^{(1)} \cdot x^{(2)}$ (interaction),

- $x^{(4)}$ denotes the treatment group $(x_i^{(4)} = 1$ for the new treatment under study and $x_i^{(4)} = 0$ for the traditional treatment)

- $T = $ survival time.

Since `survreg` by default transforms to log-land
we stay in **log-land** to begin.

We simply extend the AML two group model discussed in the **Prelude**, where we now allow for m available covariates.

Let Z denote either a standard extreme value, standard logistic, or standard normal random variable. That is, each has location $\mu = 0$ and scale $\sigma = 1$. For a given vector $\underline{x}$,

$$
\begin{aligned}
Y &= \log(T) \\
&= \widetilde{\mu} + \sigma Z \\
&= \beta_0^* + \underline{x}' \underline{\beta}^* + \sigma Z
\end{aligned}
$$

- There are only two major differences from an ordinary linear regression model for the log-transformed target variable T, $Y = \log(T)$.

  1. The distribution of the errors Z are extreme value or logistic as well as normal.

     Therefore, least-squares methods are not adequate.

  2. Our models deal with censored values, which is rarely discussed for ordinary linear regression. Recall, we obtain the MLE's by maximizing the likelihood that incorporates the censoring:

  $$
  L(\beta_0^*, \underline{\beta}^*, \sigma) = \prod_{i=1}^{n} \Big(f(y_i) \Big)^{\delta_i} \cdot \Big(S_f(y_i) \Big)^{1-\delta_i}.
  $$

- We approach the regression problem just the same as we would in ordinary least squares or even in logistic regression.

  We model the location parameter μ as a function of $\underline{x}$, linear in the $\underline{\beta}^* = (\beta_1^*, \cdots, \beta_m^*)'$, an $m \times 1$ vector of coefficients.

  In the log-normal and log-logistic cases, the location is the mean.

  In the Weibull case, the location is just location.

  In all three cases, we assume the scale σ is the same (constant) for all values of the covariates.

- The $\tilde{\mu} = \beta_0^* + \underline{x}'\underline{\beta}^*$ is called the **linear predictor**, and σ is the **scale parameter**.

- In the target variable T distribution,

$$\widetilde{\lambda} = \exp(-\widetilde{\mu}) \text{ and } \alpha = 1/\sigma.$$

- The S/R function **survreg** estimates β_0^*, $\underline{\beta}^*$, and σ.

  The **predict** function provides estimates of $\tilde{\mu}$ at specified values of the covariates.

  For example, return to the AML data, where we have one covariate "group" with two values 0 or 1. To estimate the linear predictor (**lp**) for both groups, use

```
>predict(fit,type="lp",newdata=list(group=c(0,1)),se.fit=T)
```

- The set of covariate values $\underline{x}_i$, $i = 1, \ldots, n$ are known; that is, fixed constants.

4

```
> fit<-survreg(Surv(weeks,status)~group,data=aml,
                              dist="weib")
> summary(fit)

              Value Std. Error      z         p
(Intercept)   3.180      0.241   13.22 6.89e-040
      group   0.929      0.383    2.43 1.51e-002
 Log(scale) -0.235      0.178   -1.32 1.88e-001

 Scale= 0.791

> predict(fit,type="lp",newdata=list(group=c(0,1)),
                              se.fit=T)

$fit:     [1] 3.179714        4.109055

$se.fit: [1] 0.2405564        0.2998896
```

==

The following command outputs the variance-covariance matrix $\widehat{V}$ of the estimates:

```
> fit$var
               (Intercept)         group     Log(scale)
(Intercept)   0.057867370 -0.057120613 -0.005618566
      group -0.057120613  0.146307634  0.001396857
 Log(scale) -0.005618566  0.001396857  0.031763966
```

But, internally it computes and uses $\widehat{\Sigma}$, the var-cov matrix of $\widehat{\beta}_0^*$, $\widehat{\beta}_1^*$, and $\widehat{\sigma}$.

- **The estimated linear predictor:**

$$\widehat{\mu} = \widehat{\beta}_0^* + \widehat{\beta}_1^* \text{group} = \begin{cases} \widehat{\beta}_0^*, & \text{if group} = 0 \\ \widehat{\beta}_0^* + \widehat{\beta}_1^*, & \text{if group} = 1. \end{cases}$$

- **The variances of the linear predictors:**

$$\text{var}(\widehat{\mu}) = \begin{cases} \text{var}(\widehat{\beta}_0^*), & \text{if group} = 0 \\ \text{var}(\widehat{\beta}_0^* + \widehat{\beta}_1^*), & \text{if group} = 1. \end{cases}$$

Then,

$$\text{s.e.}(\widehat{\mu}(0)) = \text{s.e.}(\widehat{\beta}_0^*) = \sqrt{(0.057867370)} = 0.2405564$$

and

$$\begin{aligned} \text{var}(\widehat{\mu}(1)) &= \text{var}(\widehat{\beta}_0^* + \widehat{\beta}_1^*) \\ &= \text{var}(\widehat{\beta}_0^*) + \text{var}(\widehat{\beta}_1^*) + 2 \cdot \text{cov}(\widehat{\beta}_0^*, \widehat{\beta}_1^*) \\ &= 0.057867370 + 0.146307634 + 2 \times (-0.057120613) \\ &= .08993378 \end{aligned}$$

and

$$\text{s.e.}(\widehat{\mu}(1)) = \sqrt{.08993378} = .2998896$$

- Back in Weibull-land; the T distribution:

$$\widehat{\lambda} = \exp(-\widehat{\mu})$$

$$= \exp(-\widehat{\beta}_0^* - \widehat{\beta}_1^* \text{group})$$

$$= \begin{cases} \exp(-\widehat{\beta}_0^*), & \text{if group} = 0 \\ \exp(-\widehat{\beta}_0^* - \widehat{\beta}_1^*), & \text{if group} = 1. \end{cases}$$

Use S/R to easily compute the Weibull parameters:

```
> fit.mu<-predict(fit,type="lp",
            newdata=list(group=c(0,1)))
> lambda.fit <- exp(-fit.mu)
> fit.mu
[1] 3.179714          4.109055
> lambda.fit
[1] 0.04159756        0.01642328
> alpha.hat <- 1/fit$scale
> alpha.hat
[1] 1.264296
```

We now have completely described both groups' distributions.

- **The recipe is same for log-normal and log-logistic models.**

Cox proportional hazards (PH) model

The hazard function:

$$h(t|\underline{x}) = h_0(t) \cdot \exp(\underline{x}'\underline{\beta}) \, ,$$

where $h_0(t)$ is an unspecified baseline hazard function free of the covariates $\underline{x}$.

• The covariates act multiplicatively on the hazard.

The hazards ratio (HR):

At two different points $\underline{x}_1, \underline{x}_2$, HR is the proportion

$$\frac{h(t|\underline{x}_1)}{h(t|\underline{x}_2)} = \frac{\exp(\underline{x}'_1\underline{\beta})}{\exp(\underline{x}'_2\underline{\beta})} = \exp\left((\underline{x}'_1 - \underline{x}'_2)\underline{\beta}\right),$$

which is constant over follow-up time t.

• This defines the **proportional hazards property**.

• Inherent in this PH property then, is the covariates $\underline{x}$ do not vary with time t.

• The exponential and Weibull are special cases when we specify the baseline hazard functions.

1. If we specify $h_0(t) = \lambda$ for all t, then this exponential.

2. If we specify $h_0(t) = \alpha\lambda^\alpha t^{\alpha-1}$, then this is Weibull.

Remarks:

1. **A statistical goal of a survival analysis is to obtain some measure of effect that will describe the relationship between a predictor variable of interest and time to failure, after adjusting for the other variables we have identified in the study and included in the model.**

• In linear regression modelling, the measure of effect is usually the regression coefficient β, the change in mean response per 1 unit change in x.

• In logistic regression, the measure of effect is an odds ratio, the log of which is β for a change of 1 unit in x.

• In survival analysis, the **measure of effect is the hazards ratio (HR)**. In the Cox PH model, this ratio is also expressed in terms of an exponential of the regression coefficient in the model.

For example, let β_1 denote the coefficient of the group covariate with group $= 1$ if received treatment and group $= 0$ if received placebo. Put treatment group in the numerator of **HR**.

A HR of 1 means that there is no effect.

A HR of 10, on the other hand, means that the treatment group has ten times the hazard of the placebo group.

Similarly, a HR of 1/10 implies that the treatment group has one-tenth the risk of failure than that of the placebo group.

2. The Weibull regression model is the only log-linear model that has the PH property.

Log-logistic Regression

The model: This log-linear model is expressed by

$$Y = \log(T)$$
$$= \beta_0^* + \underline{x}'\underline{\beta}^* + \sigma Z$$

where the error term $Z \sim$ standard logistic.

The survivor function:

$$S(t|\underline{x}) = \frac{1}{1 + (\exp(y - \beta_0^* - \underline{x}'\beta^*))^{\frac{1}{\sigma}}} \, ,$$

where $y = \log(t)$ and $\alpha = 1/\sigma$.

The odds of survival beyond time t:

$$\frac{S(t|\underline{x})}{1 - S(t|\underline{x})} = \left(\exp(y - \beta_0^* - \underline{x}'\underline{\beta}^*) \right)^{-\frac{1}{\sigma}}.$$

The $(p \times 100)$th percentile of the survival distribution:

$$t_p(\underline{x}) = \left(p/(1-p) \right)^{\sigma} \exp(\beta_0^* + \underline{x}'\underline{\beta}^*).$$

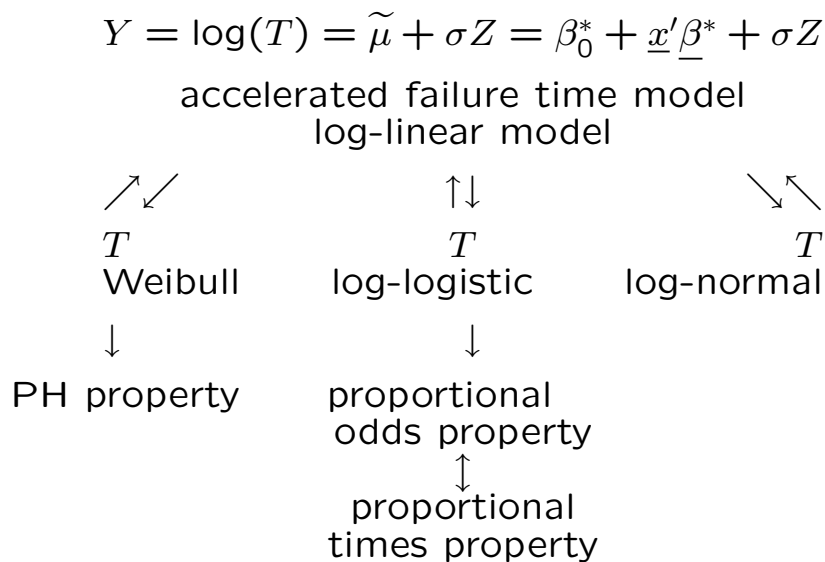**The odds-ratio (OR) of survival beyond time t evaluated at $\underline{x}_1$ and $\underline{x}_2$:**

$$\mathsf{OR}(t|\underline{x} = \underline{x}_2, \underline{x} = \underline{x}_1) = \left(\exp \left((\underline{x}_2 - \underline{x}_1)'\underline{\beta}^* \right) \right)^{\frac{1}{\sigma}} = \left(\mathsf{TR} \right)^{\frac{1}{\sigma}},$$

where TR is the times-ratio. OR is constant over follow-up time t. This model enjoys the **proportional odds property**; or equivalently, the **proportional times property**

• The reciprocal of the OR has the same functional form as the HR in the Weibull model with respect to $\underline{\beta}^*$ and σ.

• The log-logistic regression model is the only log-linear model that has the proportional odds property. Of course, we again assume the covariates $\underline{x}$ do not vary with t.

Summary

Let Z denote either a standard extreme value, standard logistic, or standard normal random variable. That is, each has location $\mu = 0$ and scale $\sigma = 1$.

$$Y = \log(T) = \widetilde{\mu} + \sigma Z = \beta_0^* + \underline{x}'\underline{\beta}^* + \sigma Z$$

accelerated failure time model
log-linear model

↗↙	↑↓	↘↖
T	T	T
Weibull	log-logistic	log-normal
↓	↓	
PH property	proportional odds property	
	↕	
	proportional times property	

• For the Weibull model, the relationship between the coefficients in the log-linear model and coefficients in modelling the hazard function is

$$\underline{\beta} = -\sigma^{-1}\underline{\beta}^* \quad \text{and} \quad \lambda = \exp(-\beta_0^*)\,.$$

• The S/R function **survreg** estimates β_0^*, $\underline{\beta}^*$, and σ.

• Fitting data to a Cox PH model is presented in detail in Chapter 5.

The Cox procedure estimates the $\underline{\beta}$ coefficients directly.

Model Building

Strategy: Use a variable selection procedure and then use the LRT for further model reduction.

1. If models are heirarchical (nested), that is, the covariates in the reduced model are a subset of the set of covariates in the full model, fit the full and reduced models via **survreg** and then use the **anova** function.

```
> anova(fit.reduced,fit.full,test="Chisq")
```

The LRT value is the value reported under `Deviance` (which should be labeled partial deviance. See Chapter 6.)

2. **Akaike's information criterion statistic for variable selection:**

$$\text{AIC} = -2 \times \log(\text{maximum likelihood}) + 2 \times (a + b),$$

where a is the number of parameters in the specified model and b is the number of one-dimensional covariates.

• The smaller the value of this statistic, the better the model.

• This statistic trades off goodness of fit (measured by the maximized log likelihood) against model complexity (measured by $a+b$).

• Can compare possible models which need not necessarily be nested nor have the same error distribution

In S/R the following sequence of code should work and is given in your Chaper 5 of the course notes:

```
> library(MASS) # Calls in a collection of library
                # functions provided by Venables
                # and Ripley


> attach(your data.frame)
> fit.start<-survreg(Surv(time,status)~X1+X2+  +Xm,
                                  dist="  ")
      # This is the initial model and has only main
      # effects, no interaction terms.
> fit.model1<- stepAIC(fit.start,~.^2)
      # Does stepwise up to two-way interaction.
> fit.model1$anova
      # Shows the stepwise model path with the
      # initial and final models
> fit.model1
    # Prints out the final model.
    # From here you try to reduce model by LRT and/or
    # one-at-a-time variable reduction by eliminating
    # the variable with the largest
    # p-value for a coefficient among those which are
    # not significant.
```

Motorette data example:

Hours to failure of motorettes are given as a function of operating temperatures 150^0C, 170^0C, 190^0C, or 220^0C. There is severe (Type I) censoring, with only 17 out of 40 motorettes failing. Note that the stress (temperature) is constant for any particular motorette over time.

Temperature	Times
150^0C	All 10 motorettes without failure at 8064 hours
170^0C	1764, 2772, 3444, 3542, 3780, 4860, 5196
	3 motorettes without failure at 5448 hours
190^0C	408, 408, 1344, 1344, 1440
	5 motorettes without failure at 1680 hours
220^0C	408, 408, 504, 504, 504
	5 motorettes without failure at 528 hours
$n = 40,$	$n_u =$ no. of uncensored times $= 17$

The **primary purpose of the experiment** was to estimate certain percentiles of the failure time distribution at a **design temperature of 130^0C.**

This is an **accelerated process**. The experiment is conducted at higher temperatures to speed up failure time. Then they make predictions at a lower temperature that would have taken them much longer to observe.

To compare our results with Nelson & Hahn and Kalbfleisch & Prentice we use the single regressor variable
$x = 1000/(273.2+\text{Temperature})$. They also omit all ten data points at temperature level of 150^0C. We entered the data into a data frame called **motorette**. It contains

time	status	temp	x
hours	1 if uncensored	^{0}C	$1000/(273.2+^0$C$)$
	0 if censored		

The **design** $x = 2.480159$.

14

$$\text{intercept only: } y \;=\; \log(t) = \beta_0^* + \sigma Z$$
$$\text{both: } y \;=\; \log(t) = \beta_0^* + \beta_1^* + \sigma Z,$$

where the distributions of Z are standard extreme (minimum) value, standard logistic, and standard normal, respectively.

Results of fitting parametric models to the Motorette data:

Model	log-likelihood		AIC	
exponential	intercept only	-155.875	$311.750 + 2(1)$	$= 313.750$
	both	-151.803	$303.606 + 2(1 + 1)$	$= 307.606$
Weibull	intercept only	-155.681	$311.363 + 2(2)$	$= 315.363$
	both	-144.345	$288.690 + 2(2 + 1)$	$= 294.690$
log-logistic	intercept only	-155.732	$311.464 + 2(2)$	$= 315.464$
	both	-144.838	$289.676 + 2(2 + 1)$	$= 295.676$
log-normal	intercept only	-155.018	$310.036 + 2(2)$	$= 314.036$
	both	-145.867	$291.735 + 2(2 + 1)$	$= 297.735$

The S/R code for computing the AIC:

```
> attach(motorette)
 # Weibull fit
> weib.fit <- survreg(Surv(time,status)~x,dist="weibull")
> weib.fit$loglik # the first component for intercept only and
                  # the second for both
[1] -155.6817  -144.3449
> -2*weib.fit$loglik # -2 times maximum log-likelihood
[1] 311.3634    288.6898
```

Nelson & Hahn applied a log-normal model, and Kalbfleisch & Prentice applied a Weibull model. Kalbfleisch & Prentice state that the Weibull model is to some extent preferable to the log-normal on account of the larger maximized log likelihood. From table, we find that the Weibull distribution provides the best fit to this data, the log-logistic distribution is a close second, and the log-normal distribution is the third.

Estimation and testing: fitting the Weibull model

Let $\widehat{\beta_0^*}$, $\widehat{\beta^{*\prime}}$, and $\widehat{\sigma}$ denote the MLE's of the parameters. Recall that the theory tells us MLE's are approximately normally distributed when the sample size n is large.

To test $H_0 : \beta_j^* = \beta_j^{*0}$, $j = 1, \ldots, m$, use

$$\frac{\widehat{\beta_j^*} - \beta_j^{*0}}{\text{s.e.}(\widehat{\beta_j^*})} \quad \overset{a}{\sim} \quad N(0, 1) \quad \text{under } H_0.$$

An approximate $(1 - \alpha) \times 100\%$ confidence interval for β_j^* is given by

$$\widehat{\beta_j^*} \pm z_{\frac{\alpha}{2}} \, \text{s.e.}(\widehat{\beta_j^*}),$$

where $z_{\frac{\alpha}{2}}$ is taken from the $N(0, 1)$ table.

At the point $\underline{x} = \underline{x}_0$, the MLE of the $(p \times 100)$th percentile of the distribution of $Y = \log(T)$ is

$$\widehat{Y}_p = \widehat{\beta_0^*} + \underline{x}_0' \widehat{\underline{\beta}^*} + \widehat{\sigma} z_p = (1, \underline{x}_0', z_p) \begin{pmatrix} \widehat{\beta_0^*} \\ \widehat{\underline{\beta}^*} \\ \widehat{\sigma} \end{pmatrix},$$

where z_p is the $(p \times 100)$th percentile of the error distribution, which, in this case, is standard extreme value.

This is a linear combination of the estimates. Need to consider the covariances when computing the variance of $\widehat{Y}_p$. Internally, survreg computes $\widehat{\Sigma}$, the estimated variance-covariance matrix of $\widehat{\beta_0^*}$, $\widehat{\beta_1^*}$, and $\widehat{\sigma}$.

An approximate $(1-\alpha)\times 100\%$ confidence interval for the $(p\times 100)$th percentile of the log-failure time distribution is given by

$$\widehat{Y}_p \pm \mathsf{z}_{\frac{\alpha}{2}}\, \mathsf{s.e.}(\widehat{Y}_p),$$

where $\mathsf{z}_{\frac{\alpha}{2}}$ is taken from the $N(0,1)$ table.

The MLE of t_p is $\exp(\widehat{Y}_p)$. To obtain confidence limits for t_p, take the exponential of the endpoints of the above confidence interval.

Recall, the function `predict`, a companion function to `survreg`, conveniently reports both the quantiles in time and the uquantiles in log(time) along with their respective s.e.'s.

Doing the analysis using S/R:

```
> fit.weib <- survreg(Surv(time,status)~x,data=motorette)
> summary(fit.weib)

              Value Std. Error      z          p
(Intercept) -11.89       1.966  -6.05 1.45e-009
          x   9.04       0.906   9.98 1.94e-023
 Log(scale)  -1.02       0.220  -4.63 3.72e-006


Scale= 0.361

> predict(fit.weib,type="uquantile",newdata=list(x = 2.480159),
                se.fit=T,p=c(0.15, 0.5, 0.85))
$fit:    [1]  9.868867  10.392887    10.756643


$se.fit: [1] 0.3444804  0.3026464    0.2973887

> predict(fit.weib,type="quantile",newdata=list(x=2.480159),
                se.fit = T, p = c(0.15, 0.5, 0.85))
$fit:    [1]  19319.44    32626.72    46940.83


$se.fit: [1]  6655.168    9874.361    13959.673

> mutilde.hat <- predict(fit.weib,type="lp",newdata=list(x=2.480159),
                    se.fit=T)
> lambdatilde.hat <- exp(-mutilde.hat$fit)
> mutilde.hat
$fit:    [1] 10.5253


$se.fit: [1] 0.2982668

> lambdatilde.hat
[1] 0.00002684848
> fit.weib$scale
[1] 0.3612814
```
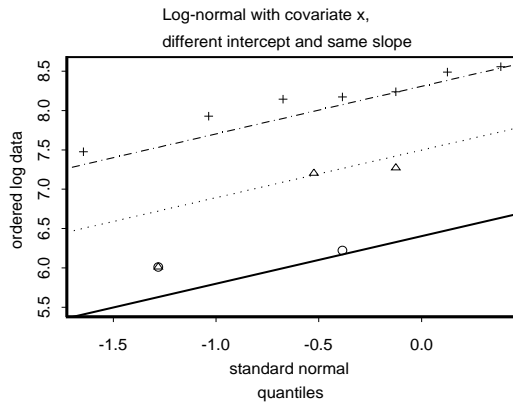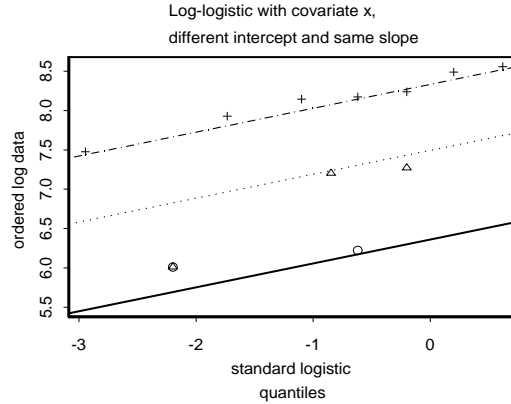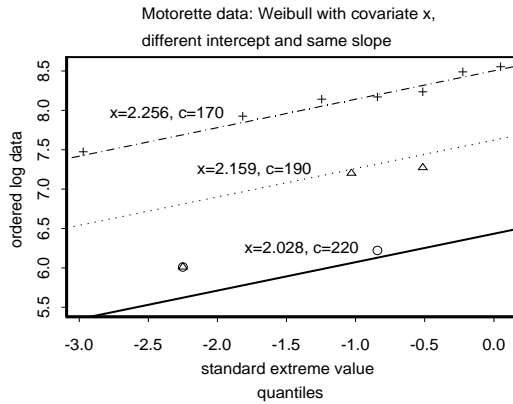
```
> alpha.hat <- 1/fit.weib$scale
> alpha.hat
[1] 2.767925
> pweibull(25000,alpha.hat,1/lambdatilde.hat)
     # Computes the
     # estimated probability that a motorette failure time
     # is less than or equal to 25,000 hours.

 [1] 0.2783054 # estimated probability

> Shatq <- 1 - 0.2783054
     # survival probability at 25,000
     # hours. About 72% of motorettes are still working
     # after 25,000 hours at x=2.480159; i.e., the design
     # temperature of 130 degrees Celsius.
```

Motorette data: Weibull with covariate x, different intercept and same slope

Log-logistic with covariate x, different intercept and same slope

Log-normal with covariate x, different intercept and same slope

Summary of some results:

- From `summary(weib.fit)`, we learn that $\widehat{\sigma} = .36128$, and $\widehat{\widetilde{\mu}} = -\log(\widehat{\widetilde{\lambda}}) = \widehat{\beta}_0^* + \widehat{\beta}_1^* x = -11.89 + 9.04x$.

  The $\widehat{\alpha} = 1/.36128 = 2.768$ and $\widehat{\widetilde{\lambda}} = \exp(11.89 - 9.04 \times 2.480159) = 0.0000268$ at $x = 2.480159$.

  Note also that both the intercept and covariate x are highly significant with p-values 1.45×10^{-9} and 1.94×10^{-23}, respectively.

- The estimated hazard function is

$$\widehat{h}(t|x) = \frac{1}{\widehat{\sigma}} \cdot t^{\frac{1}{\widehat{\sigma}} - 1} \cdot (\exp(-\widehat{\mu}))^{\frac{1}{\widehat{\sigma}}}$$

  and the estimated survivor function is

$$\widehat{S}(t|x) = \exp\left\{ - \left(\exp(-\widehat{\mu})t \right)^{\frac{1}{\widehat{\sigma}}} \right\}.$$

- The point estimate of β_1, $\widehat{\beta}_1$, is $-\widehat{\sigma}^{-1}\widehat{\beta}_1^*$. A 95% C.I. for β_1 based on the delta method is given by $[-37.84342, -12.29594]$. Whereas the one based on the common approach is given by

$$[-\widehat{\sigma}^{-1}(10.82), -\widehat{\sigma}^{-1}(7.26)] = [-29.92, -20.09],$$

  where $\widehat{\sigma} = .3605949$ and the 95% C.I. for β_1^* is $[7.26, 10.81] = [9.04 - 1.96 \times .906, \ 9.04 + 1.96 \times .906]$. It is clear that the latter interval is much shorter than the former as it ignores the variability of $\widehat{\sigma}$.

- At $x = 2.480159$, the design temperature of 130^0C, the estimated 15th, 50th, and 85th percentiles in log(hours) and hours, respectively based on uquantile and quantile, along with their corresponding 90% C.I.'s in hours are reported in the following table.

type	percentile	Estimate	Std.Err	90% LCL	90% UCL
uquantile	15	9.8689	0.3445	10962.07	34048.36
	50	10.3929	0.3026	19831.64	53677.02
	85	10.7566	0.2974	28780.08	76561.33
quantile	15	19319.44	6655.168	9937.17	37560.17
	50	32626.76	9874.36	19668.76	54121.65
	85	46940.83	13959.67	28636.93	76944.21

  The 90% C.I.'s based on uquantile, exp(estimate $\pm$ 1.645 $\times$ std.err), are shorter than those based on quantile at each x value.

- At the design temperature 130^0C, by 25,000 hours about 28% of the motorettes have failed. That is, after 25,000 hours, about 72% are still working.

- As $\hat{\alpha} = \frac{1}{\hat{\sigma}} = \frac{1}{.3605949} = 2.773195$, then for fixed x the hazard function increases as time increases.

- The estimated coefficient $\hat{\beta}_1 = -\frac{1}{\hat{\sigma}}\hat{\beta}_1^* = -\frac{1}{.3605949}(9.04) = -25.06968 < 0$. Thus, for time fixed, as x increases, the hazard decreases and survival increases.

- For $x_1 < x_2$,

$$\frac{h(t|x_2)}{h(t|x_1)} = \exp((x_2 - x_1)(-25.06968)).$$

For example, for $x = 2.1$ and 2.2,

$$\frac{h(t|2.2)}{h(t|2.1)} = \exp(.1(-25.06968)) = .08151502.$$

For .1 unit increase in x, the hazard becomes about 8.2% of the hazard before the increase. In terms of Celsius temperature, for 21.645 degree decrease from 202.9905^0C to 181.3455^0C, the hazard becomes about 8.2% of the hazard before the decrease.

- The Q-Q plots show that the Weibull fit looks slightly better than the log-logistic fit at the temperature 170^0C, but overall they are the same. On the other hand, the Weibull fit looks noticeably better than the log-normal fit at the temperature 170^0C and is about the same at the other two temperatures. This result coincides with our finding from AIC; that is, among these three accelerated failure time models, the Weibull best describes the motorette data.

Chapter 5

A Prognostic Factor Analysis based on Cox PH Model

Example: CNS lymphoma data

The data result from an observational clinical study conducted at Oregon Health Sciences University (OHSU). Fifty-eight non-AIDS patients with central nervous system (CNS) lymphoma were treated at OHSU from January 1982 through March of 1992. Group 1 patients (n=19) received cranial radiation prior to referral for blood-brain barrier disruption (BBBD) chemotherapy treatment; Group 0 (n=39) received, as their initial treatment, the BBBD chemotherapy treatment. Radiographic tumor response and survival were evaluated.
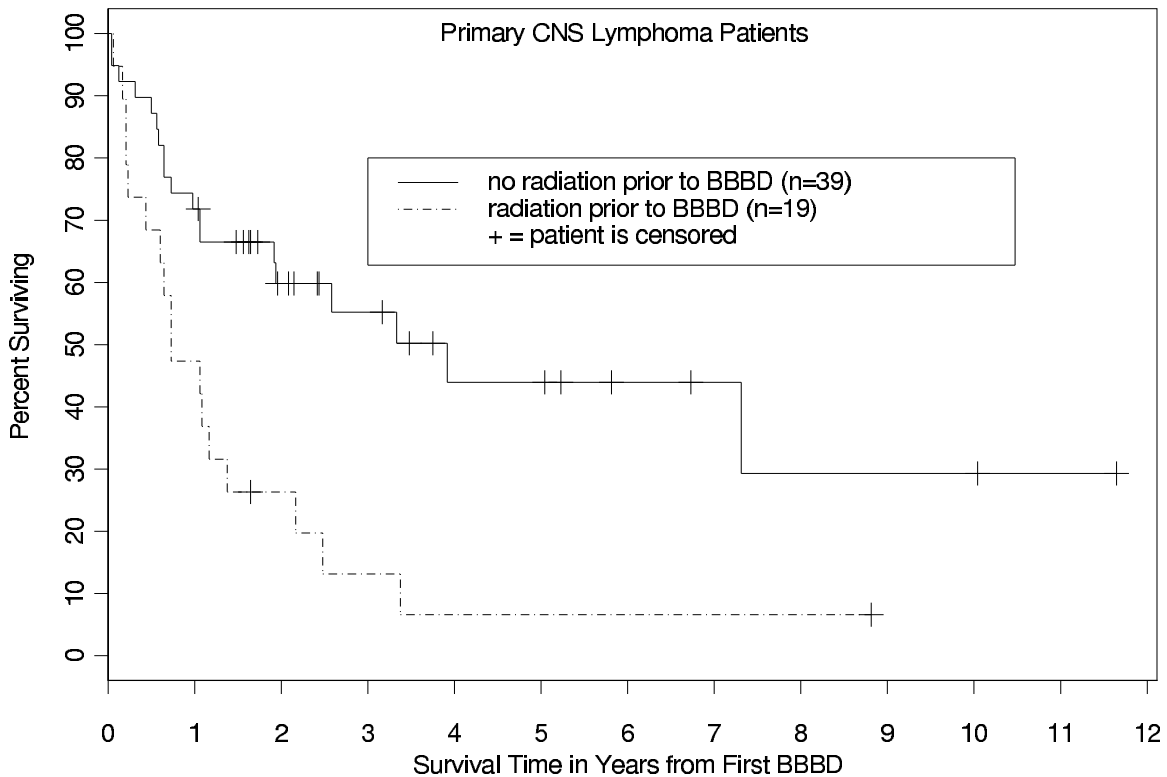
The primary endpoint of interest here is survival time (in years) from first BBBD to death (B3TODEATH). Some questions of interest are:

1. Is there a difference in survival between the two groups?

2. Do any subsets of available covariates help explain this survival time? For example, does age at time of first treatment and/or gender increase or decrease the hazard of death; hence, decrease or increase the probability of survival; and hence, decrease or increase mean or median survival time?

3. Is there a dependence of the difference in survival between the groups on any subset of the available covariates?

```
> cns2.fit0 <- survfit(Surv(B3TODEATH,STATUS)~GROUP,data=cns2,
      type="kaplan-meier")
> plot(cns2.fit0,type="l",....)
```


Primary CNS Lymphoma Patients

no radiation prior to BBBD (n=39)
radiation prior to BBBD (n=19)
+ = patient is censored

Percent Surviving

Survival Time in Years from First BBBD

```
> survdiff(Surv(B3TODEATH,STATUS)~GROUP,data=cns2)
          N Observed Expected (O-E)^2/E (O-E)^2/V
 GROUP=0 39       19    26.91      2.32      9.52
 GROUP=1 19       17     9.09      6.87      9.52
 Chisq= 9.5  on 1 degrees of freedom, p= 0.00203

> cns2.fit0


          n events mean se(mean) median 0.95LCL 0.95UCL
 GROUP=0 39     19 5.33    0.973  3.917   1.917      NA
 GROUP=1 19     17 1.57    0.513  0.729   0.604    2.48
```

In the Cox PH model, recall the HR at two different points $\underline{x}_1$ and $\underline{x}_2$, is the proportion

$$
\begin{aligned}
\frac{h(t|\underline{x}_1)}{h(t|\underline{x}_2)} &= \frac{\exp(\underline{x}_1' \underline{\beta})}{\exp(\underline{x}_2' \underline{\beta})} \\
&= \frac{\exp\left(\beta_1 x_1^{(1)}\right) \times \exp\left(\beta_2 x_1^{(2)}\right) \times \cdots \times \exp\left(\beta_m x_1^{(m)}\right)}{\exp\left(\beta_1 x_2^{(1)}\right) \times \exp\left(\beta_2 x_2^{(2)}\right) \times \cdots \times \exp\left(\beta_m x_2^{(m)}\right)}
\end{aligned}
$$

which is constant over follow-up time t.

Through the **partial likelihood** we obtain estimates of the coefficients $\underline{\beta}$ without regard to the baseline hazard $h_0(t)$. The partial likelihood is presented in Chapter 6.

Note that in the parametric regression setting of Chapter 4, we specify the form of this function since we must specify a distribution for the target variable T.

Remember the hazard function completely specifies the distribution of T; but the power of the PH model is that it provides a fairly wide family of distributions by allowing the baseline hazard $h_0(t)$ to be arbitrary.

The S/R function `coxph` implements Cox's partial likelihood function.

AIC procedure for variable selection

$$\text{AIC} = -2 \times \log(\text{maximum likelihood}) + 2 \times b,$$

where b is the number of β coefficients in each model under consideration. The **maximum likelihood is replaced by the maximum partial likelihood**. The smaller the AIC value the better the model is.

We apply an automated model selection procedure via an S function `stepAIC` included in MASS, a collection of functions and data sets from *Modern Applied Statistics with S* by Venables and Ripley (2002). Otherwise, it would be too tedious because of many steps involved.

We illustrate how to use `stepAIC` together with LRT to select a best model.

The estimates from fitting a Cox PH model are interpreted as follows:

- A positive coefficient increases the risk and thus decreases the expected (average) survival time.

- A negative coefficient decreases the risk and thus increases the expected survival time.

- The ratio of the estimated risk functions for the two groups can be used to examine the likelihood of Group 0's (no prior radiation) survival time being longer than Group 1's (with prior radiation).

Step I: `stepAIC` **to select the best model according to AIC statistic**

```
> library(MASS) # Call in a collection of library functions
                # provided by Venables and Ripley
> attach(cns2)
> cns2.coxint<-coxph(Surv(B3TODEATH,STATUS)~KPS.PRE.+GROUP+SEX+
  AGE60+LESSING+LESDEEP+factor(LESSUP)+factor(PROC)+CHEMOPRIOR)
    # Initial model
> cns2.coxint1 <- stepAIC(cns2.coxint,~.^2)
                    # Up to two-way interaction
> cns2.coxint1$anova # Shows stepwise model path with the
                     # initial and final models
```

Stepwise model path for two-way interaction model on the CNS lymphoma data

Step	Df	AIC
		246.0864
+ SEX:AGE60	1	239.3337
- factor(PROC)	2	236.7472
- LESDEEP	1	234.7764
- factor(LESSUP)	2	233.1464
+ AGE60:LESSING	1	232.8460
+ GROUP:AGE60	1	232.6511

27

Step II: LRT to further reduce

```
> cns2.coxint1 # Check which variable has a
               # moderately large p-value

                   coef exp(coef) se(coef)      z       p
      KPS.PRE. -0.0471    0.9540    0.014 -3.362 0.00077
         GROUP  2.0139    7.4924    0.707  2.850 0.00440
           SEX -3.3088    0.0366    0.886 -3.735 0.00019
         AGE60 -0.4037    0.6679    0.686 -0.588 0.56000
       LESSING  1.6470    5.1916    0.670  2.456 0.01400
    CHEMOPRIOR  1.0101    2.7460    0.539  1.876 0.06100
     SEX:AGE60  2.8667   17.5789    0.921  3.113 0.00190
 AGE60:LESSING -1.5860    0.2048    0.838 -1.891 0.05900
   GROUP:AGE60 -1.2575    0.2844    0.838 -1.500 0.13000
```

In statistical modelling, an important principle is that an interaction term should only be included in a model when the corresponding main effects are also present.

We now see if we can eliminate the variable AGE60 and its interaction terms with other variables.

Here the LRT is constructed on the partial likelihood function rather than the full likelihood function.

```
> cns2.coxint2 <- coxph(Surv(B3TODEATH,STATUS)~KPS.PRE.+GROUP
        +SEX+LESSING+CHEMOPRIOR) # Without AGE60 and its
                                 # interaction terms
> -2*cns2.coxint2$loglik[2] + 2*cns2.coxint1$loglik[2]
[1] 13.42442
> 1 - pchisq(13.42442,4)
[1] 0.009377846 # Retain the model selected by stepAIC
```

Now begin the process of one variable at a time reduction.

- The variable GROUP:AGE60 has a moderately large
  p-value $= .130$. Delete it.

```
> cns2.coxint3 # Check which variable has a
              # moderately large p-value
```

```
                  coef exp(coef) se(coef)     z      p
   KPS.PRE. -0.0436     0.9573   0.0134 -3.25 0.0011
      GROUP  1.1276     3.0884   0.4351  2.59 0.0096
        SEX -2.7520     0.0638   0.7613 -3.61 0.0003
      AGE60 -0.9209     0.3982   0.5991 -1.54 0.1200
    LESSING  1.3609     3.8998   0.6333  2.15 0.0320
 CHEMOPRIOR  0.8670     2.3797   0.5260  1.65 0.0990
   SEX:AGE60  2.4562    11.6607   0.8788  2.79 0.0052
AGE60:LESSING -1.2310     0.2920   0.8059 -1.53 0.1300
```

- As AGE60:LESSING has a moderately large
  p-value $= .130$, we remove it.

```
> cns2.coxint4 # Check which variable has a
              # moderately large p-value
```

```
                 coef exp(coef) se(coef)     z       p
  KPS.PRE. -0.0371     0.9636   0.0124 -3.00 0.00270
     GROUP  1.1524     3.1658   0.4331  2.66 0.00780
       SEX -2.5965     0.0745   0.7648 -3.40 0.00069
     AGE60 -1.3799     0.2516   0.5129 -2.69 0.00710
   LESSING  0.5709     1.7699   0.4037  1.41 0.16000
CHEMOPRIOR  0.8555     2.3526   0.5179  1.65 0.09900
 SEX:AGE60  2.3480    10.4643   0.8765  2.68 0.00740
```

- We delete the term LESSING as it has a moderately large
  p-value $= .160$.

```
> cns2.coxint5 # Check which variable has a
              # moderately large p-value


             coef exp(coef) se(coef)     z       p
   KPS.PRE. -0.0402    0.9606   0.0121 -3.31 0.00093
     GROUP  0.9695     2.6366   0.4091  2.37 0.01800
       SEX -2.4742     0.0842   0.7676 -3.22 0.00130
     AGE60 -1.1109     0.3293   0.4729 -2.35 0.01900
 CHEMOPRIOR  0.7953    2.2152   0.5105  1.56 0.12000
  SEX:AGE60  2.1844    8.8856   0.8713  2.51 0.01200
```

• We eliminate the variable CHEMOPRIOR as it has a moderately large p-value $= .120$.

```
> cns2.coxint6 # Check which variable has a
              # moderately large p-value


             coef exp(coef) se(coef)     z      p
 KPS.PRE. -0.0307     0.970   0.0102 -2.99 0.0028
    GROUP  1.1592     3.187   0.3794  3.06 0.0022
      SEX -2.1113     0.121   0.7011 -3.01 0.0026
    AGE60 -1.0538     0.349   0.4572 -2.30 0.0210
SEX:AGE60  2.1400     8.500   0.8540  2.51 0.0120
```

• We finally stop here and retain these five variables: KPS.PRE., GROUP, SEX, AGE60, and SEX:AGE60.

• However, it is important to compare this model to the model chosen by `stepAIC` in Step I as we have not compared them.

```
> -2*cns2.coxint6$loglik[2] + 2*cns2.coxint1$loglik[2]
[1] 8.843838
> 1 - pchisq(8.843838,4)
[1] 0.06512354 # Selects the reduced model
```

• The p-value based on LRT is between .05 and .1. So we select the reduced model with caution.

• The following output is based on the model with KPS.PRE., GROUP, SEX, AGE60, and SEX:AGE60. It shows that the three tests − LRT, Wald, and efficient score test − indicate there is an overall significant relationship between this set of covariates and survival time. That is, they are explaining a significant portion of the variation.

```
> summary(cns2.coxint6)

 Likelihood ratio test= 27.6  on 5 df,   p=0.0000431
 Wald test             = 24.6  on 5 df,   p=0.000164
 Score (logrank) test = 28.5  on 5 df,   p=0.0000296
```

Remarks:

1. The model selection procedure may well depend on the purpose of the study. In some studies there may be a few variables of special interest. In this case, we can still use Step I and Step II. In Step I we select the best set of variables according to the smallest AIC statistic. If this set includes all the variables of special interest, then in Step II we have only to see if we can further reduce the model. Otherwise, add to the selected model the unselected variables of special interest and go through Step II.

2. It is important to include interaction terms in model selection procedures unless researchers have compelling reasons why they do not need them. We could end up with a quite different model when only main effects models are considered.

Discussion

- KPS.PRE., GROUP, SEX, AGE60, and SEX:AGE60 appear to have a significant effect on survival duration. Here it is confirmed again that there is a significant difference between the two groups' (0=no prior radiation,1=prior radiation) survival curves.

- The estimated coefficient for KPS.PRE. is $-.0307$ with p-value 0.0028. Hence, fixing other covariates, patients with high KPS.PRE. scores have a decreased hazard, and, hence, have longer expected survival time than those with low KPS.PRE. scores.

- The estimated coefficient for GROUP is 1.1592 with p-value 0.0022. Hence, with other covariates fixed, patients with radiation prior to first BBBD have an increased hazard, and, hence, have shorter expected survival time than those in Group 0.

- Fixing other covariates, the hazard ratio between Group 1 and Group 0 is

$$\frac{\exp(1.1592)}{\exp(0)} = 3.187.$$

This means that, with other covariates fixed, patients with radiation prior to first BBBD are 3.187 times more likely than those without to have shorter survival.

- Fixing other covariates, if a patient in Group 1 has 10 units larger KPS.PRE. score than a patient in Group 0, the ratio of hazard functions is

$$\frac{\exp(1.1592)\exp(-0.0307 \times (k+10))}{\exp(0)\exp(-.0307 \times k)} = \frac{\exp(1.1592)\exp(-0.0307 \times 10)}{\exp(0)}$$
$$= 3.187 \times 0.7357 = 2.345,$$

where k is an arbitrary number. This means that fixing other covariates, a patient in Group 1 with 10 units larger KPS.PRE. score than a patient in Group 0 is 2.34 times more likely to have shorter survival.

In summary, fixing other covariates, whether a patient gets radiation therapy prior to first BBBD is more important than how large his/her KPS.PRE. score is.

- There is significant interaction between AGE60 and SEX. The estimated coefficient for SEX:AGE60 is 2.1400 with p-value 0.0120. Fixing other covariates, a male patient who is younger than 60 years old has 34.86% the risk a male older than 60 years old has of succumbing to the disease, where

$$\frac{\exp(-2.113 \times 0 - 1.0538 \times 1 + 2.14 \times 0)}{\exp(-2.113 \times 0 - 1.0538 \times 0 + 2.14 \times 0)} = \exp(-1.0538) = .3486.$$

Whereas, fixing other covariates, a female patient who is younger than 60 years old has 2.963 times the risk a female older than 60 years old has of succumbing to the disease, where

$$\frac{\exp(-2.113 \times 1 - 1.0538 \times 1 + 2.14 \times 1)}{\exp(-2.113 \times 1 - 1.0538 \times 0 + 2.14 \times 0)} = \exp(1.0862) = 2.963.$$

We plot the interaction between SEX and AGE60 based on the means computed using `survfit` for the response and AGE60, fixing female and male separately. It shows a clear pattern of interaction, which supports the prior numeric results using Cox model cns2.coxint6.

Interaction between SEX and AGE60
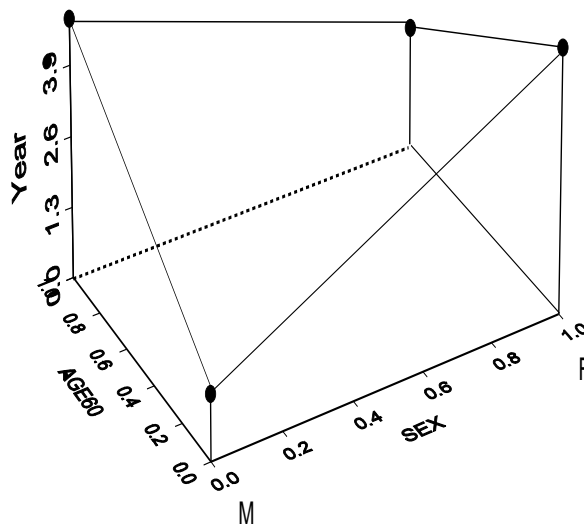
To produce the panel figure, we first fit the data to the model

```
> cox.fit <- coxph(Surv(B3TODEATH,STATUS)~KPS.PRE.+GROUP+
                        strata(factor(SEX),factor(AGE60)))
```

which adjusts for the GROUP and KPS.PRE. effects. We then set GROUP = 1, KPS.PRE. = 80 and obtain the summary of the adjusted quantiles and means using `survfit` as follows:

```
> survfit(cox.fit,data.frame(GROUP=1,KPS.PRE.=80))
> summary(survfit(cox.fit,data.frame(GROUP=1,KPS.PRE.=80)))
```

The panel displays both ordinal and disordinal interactions. The survival curve for females who are older than 60 years never steps down below 0.50. In order to produce the median plot, we set the median survival time since 1st BBBD for this stratum at 1.375 years, which is the .368-quantile.

Interaction between SEX and AGE60 adjusted for KPS.PRE. and GROUP via `coxph` and then evaluated at GROUP = 1 and KPS.PRE. = 80
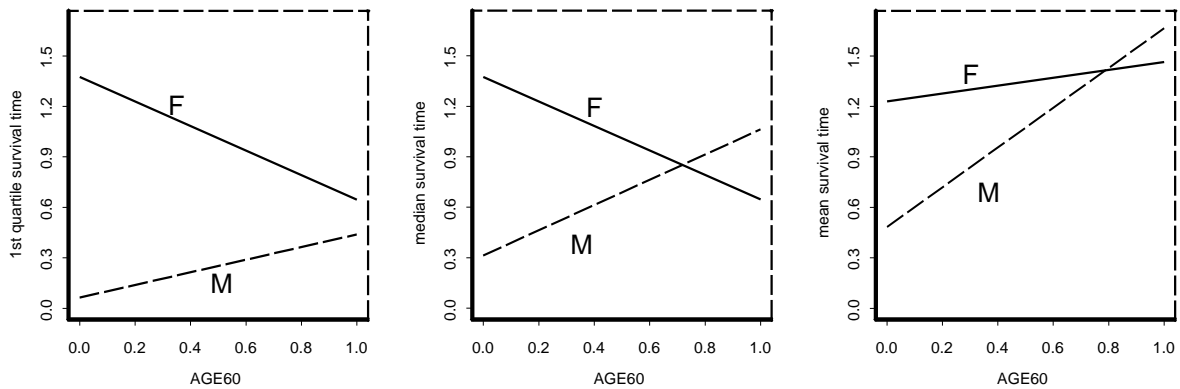
Crossing Hazards? What does one do?

Example: Extracted from Kleinbaum (1996).

A study in which cancer patients are randomized to either surgery or radiation therapy without surgery is considered. We have a $(0, 1)$ exposure variable E denoting surgery status, with 0 if a patient receives surgery and 1 if not (i.e., receives radiation). Suppose further that this exposure variable is the only variable of interest.

Is the Cox PH model appropriate? To answer this note that when a patient undergoes serious surgery, as when removing a cancerous tumor, there is usually a high risk for complications from surgery or perhaps even death early in the recovery process, and once the patient gets past this early critical period, the benefits of surgery, if any, can be observed.

In this study that compares surgery to no surgery, we might expect

to see hazard functions for each group as follows:
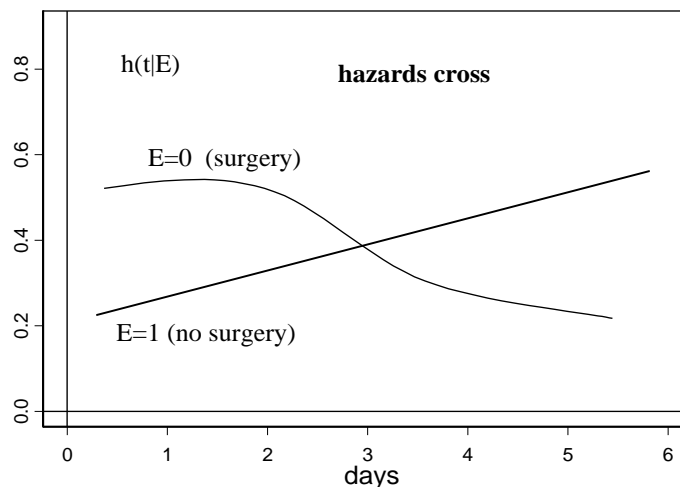
Before 2 years, HR(1|0) < 1, whereas later, HR(1|0) > 1.

The PH assumption is violated, since HR must be constant over the follow-up time.

If the Cox PH model is inappropriate, there are several options available for the analysis:

- analyze by **stratifying** on the exposure variable; that is, do not fit any regression model, and, instead obtain the Kaplan-Meier curve for each group separately;

- to adjust for other significant factor effects, use Cox model stratified on exposure variable E.

  > coxph(Surv(time,status)~X1+X2+···+strata(E))

- start the analysis at three days, and use a Cox PH model on three-day survivors;

- fit a Cox PH model for less than three days and a different Cox PH model for greater than three days to get two different hazard ratio estimates, one for each of these two time periods;

- fit a Cox PH model that includes a time-dependent variable which measures the interaction of exposure with time. This model is called an **extended Cox model** and is presented in Chapter 7 of our book.

- use the **censored regression quantile** approach, presented in Chapter 8 of our book, which allows crossover effects. This approach is nonparametric and is free of the PH assumption for its validity.

Chapter 6

Model Checking: Data Diagnostics

A very brief summary for the Cox PH model

Let $t_{(1)}, \ldots, t_{(r)}$ denote the $r \leq n$ ordered (uncensored) death times, so that $t_{(j)}$ is the jth ordered death time. Let $\underline{x}_{(j)}$ denote the vector of covariates associated with the individual who dies at $t_{(j)}$.

The Cox partial likelihood for this j-th individual is given by

$$L_j(\underline{\beta}) = P\{\text{individual with } \underline{x}_{(j)} \text{ dies at } t_{(j)} | \text{one death in } \mathcal{R}(t_{(j)}) \text{ at } t_{(j)}\}$$

$$= \frac{\exp(\underline{x}'_{(j)}\underline{\beta})}{\sum_{l \in \mathcal{R}(t_{(j)})} \exp(\underline{x}'_l\underline{\beta})}.$$

Do you see the connection here with the form of HR?

Cox's partial likelihood function:

$$L_c(\underline{\beta}) = \prod_{j=1}^{r} L_j(\underline{\beta}) = \prod_{j=1}^{r} \frac{\exp(\underline{x}'_{(j)}\underline{\beta})}{\sum_{l \in \mathcal{R}(t_{(j)})} \exp(\underline{x}'_l\underline{\beta})}.$$

An equivalent expression for the partial likelihood function in terms of all n observed times:

$$L_c(\underline{\beta}) = \prod_{i=1}^{n} \left(\frac{\exp(\underline{x}'_i\underline{\beta})}{\sum_{l \in \mathcal{R}(y_i)} \exp(\underline{x}'_l\underline{\beta})} \right)^{\delta_i}.$$

Remarks:

1. Cox's estimates maximize the log-partial likelihood.

2. To analyze the effect of covariates, there is no need to estimate the nuisance parameter $h_0(t)$, the baseline hazard function.

3. Cox argues that most of the relevant information about the coefficients β for regression with censored data is contained in this partial likelihood.

4. This partial likelihood is not a true likelihood in that it does not integrate out to 1 over $\{0, 1\}^n \times \Re_+{}^n$.

5. Censored individuals do not contribute to the numerator of each factor. But they do enter into the summation over the risk sets at death times that occur before a censored time.

6. This partial likelihood depends only on the ranking of the death times, since this determines the risk set at each death time. So, the estimates are a function of the rank order of the death times only. This is why we say this is a nonparametric method.

   Ranks are distribution free.

An heuristic derivation of Cox's partial likelihood function:

Let t^* denote a time at which a death has occurred. Let $\mathcal{R}(t^*)$ be the risk set at time t^*; that is, the indices of individuals who are alive and not censored just before t^*. First,

$$P\{\text{one death in } [t^*, t^* + \triangle t^*) \mid \mathcal{R}(t^*)\}$$

$$= \sum_{l \in \mathcal{R}(t^*)} P\{T_l \in [t^*, t^* + \triangle t^*) \mid T_l \geq t^*\}$$

$$\approx \sum_{l \in \mathcal{R}(t^*)} h(t^* | \underline{x}_l) \triangle t^*$$

$$= \sum_{l \in \mathcal{R}(t^*)} h_0(t^*) \cdot \exp(\underline{x}_l' \underline{\beta}) \triangle t^*.$$

Thus, if we let $P\{\text{one death at } t^* \mid \mathcal{R}(t^*)\}$ denote the

$$\sum_{l \in \mathcal{R}(t^*)} P(T_l = t^* | T_l \geq t^*) \, ,$$

then we have

$$P\{\text{one death at } t^* \mid \mathcal{R}(t^*)\} = \sum_{l \in \mathcal{R}(t^*)} h_0(t^*) \cdot \exp(\underline{x}_l' \underline{\beta}).$$

Now, let $t_{(1)}, \ldots, t_{(r)}$ denote the $r \le n$ ordered (uncensored) death times, so that $t_{(j)}$ is the jth ordered death time. Let $\underline{x}_{(j)}$ denote the vector of covariates associated with the individual who dies at $t_{(j)}$. Then, for each j, we have

$$
\begin{aligned}
L_j(\underline{\beta}) \quad &= \quad P\{\text{individual with } \underline{x}_{(j)} \text{ dies at } t_{(j)} \mid \text{one death in } \mathcal{R}(t_{(j)}) \text{ at } t_{(j)}\} \\[2mm]
&= \quad \frac{P\{\text{individual with } \underline{x}_{(j)} \text{ dies at } t_{(j)} \mid \text{individual in } \mathcal{R}(t_{(j)})\}}{P\{\text{one death at } t_{(j)} \mid \mathcal{R}(t_{(j)})\}} \\[2mm]
&= \quad \frac{h_0(t_{(j)}) \cdot \exp(\underline{x}'_{(j)}\underline{\beta})}{\sum_{l \in \mathcal{R}(t_{(j)})} h_0(t_{(j)}) \cdot \exp(\underline{x}'_l \underline{\beta})} \\[2mm]
&= \quad \frac{\exp(\underline{x}'_{(j)}\underline{\beta})}{\sum_{l \in \mathcal{R}(t_{(j)})} \exp(\underline{x}'_l \underline{\beta})}.
\end{aligned}
$$

Six most popular residual plots to examine:

1. Cox-Snell residuals for assessing the overall fit of a PH model

2. Martingale residuals for identifying the best functional form of a covariate

3. Deviance residuals to detect possible outliers

4. Schoenfeld residuals to examine fit and detect outlying covariate values
   - also called partial residuals and score residuals

5. Scaled Schoenfeld residuals and Grambsch and Therneau's test for PH assumption

6. dfbetas to assess influence of each observation