In the previous sections, we looked at non-parametric methods.

However, in most studies there are explanatory variables for which we want to look at their influence on the time till event.

The most commonly used classes of regression models in survival data are

- Cox's regression model or proportional hazard model

$$\lambda(t|\mathbf{X}) = \lambda_0(t)g(\beta_1 X_1 + \ldots + \beta_p X_p)$$

with $g \geq 0$ a known function, mostly $g(u) = e^u$.

- Accelerated failure time model (AFT)

$$\log(T) = \mu + \beta_1 X_1 + \ldots + \beta_p X_p + \sigma W$$

with W a (parametric) error-distribution.

Here, we study the Cox's regression model or proportional hazard model, introduced by Cox (1972).

In this model, the conditional hazard of an individual, given the covariate values $X_1, \ldots, X_p$, is defined as

$$\lambda(t|\mathbf{X}) = \lambda_0(t)e^{\beta_1 X_1 + \ldots + \beta_p X_p} = \lambda_0(t)e^{\beta^t \mathbf{X}}.$$

where $\lambda_0(t)$ is called the baseline hazard.

The strength of this model is that $\lambda_0(t)$ is left unspecified (unknown function). It represents the hazard of an individual with covariates equal to zero.

We call this a semi-parametric regression model.

This model is often also called proportional hazard model.

Consider two individuals with covariates $\mathbf{X}$ and $\mathbf{X}^*$, the ratio of their hazards is

$$HR(t) = \frac{\lambda(t|\mathbf{X})}{\lambda(t|\mathbf{X}^*)} = \frac{\lambda_0(t)e^{\beta^t\mathbf{X}}}{\lambda_0(t)e^{\beta^t\mathbf{X}^*}} = e^{\sum_{i=1}^{p}\beta_i(X_i - X_i^*)}$$

which is a constant over time.

This quantity is called the hazard ratio and compares the hazard of having an event with covariate value $\mathbf{X}$ to the hazard of having an event with covariate value $\mathbf{X}^*$.

Often, the ratio is (inappropriately) called a *relative risk*.

If X is an indicator for gender: Male ($X = 0$) vs. Female ($X = 1$), we get that

$$HR(t) = \frac{\lambda_F(t)}{\lambda_M(t)} = e^{\beta_1}.$$

In this model, we get for Males: $\lambda_M(t) = \lambda_0(t)$ and for Females: $\lambda_F(t) = \lambda_0(t)e^{\beta_1}$.

Using a different coding for gender: Male ($X = 1$) vs. Female ($X = 0$), we get for Males: $\lambda_M(t) = \lambda_0(t)e^{\beta_2}$ and for Females: $\lambda_F(t) = \lambda_0(t)$.

Hence, we get the same results for $HR(t) = e^{\beta_1} = e^{-\beta_2}$ but selecting the baseline is important.

For a continuous variable X, we note that the hazard ratio of level $x+1$ versus level x is given by

$$\frac{\lambda(t|X=x+1)}{\lambda(t|X=x)} = \frac{\lambda_0(t)e^{\beta(x+1)}}{\lambda_0(t)e^{\beta x}} = e^{\beta}.$$

Hence, in Cox's regression model

$$\lambda(t|X) = \lambda_0(t)e^{\beta X}$$

So e^{β} describes the proportional change of the hazard due to the increase of X by one unit.

We note that

$$\beta > 0 \Rightarrow \text{hazard increases.}$$

$$\beta < 0 \Rightarrow \text{hazard decreases.}$$

Example: NSCLC

- Laudanski et al., Eur Respir J (2001).
- In this study, we had 102 patients who were operated from long cancer.
- The severity of the cancer was expressed in three TNM (Tumor, Nodes, Metastasis) categories: I, II, IIIa.
- The expression of the P53 protein was found from tumor biopsies.

Let's take $X = 0, 1, 2$ for, respectively, TNM = I, II, IIIa.

Hence, we get

- hazard function for TNM I: $\lambda(t|X = 0) = \lambda_0(t)$.
- hazard function for TNM II: $\lambda(t|X = 1) = \lambda_0(t)e^{\beta}$.
- hazard function for TNM IIIa: $\lambda(t|X = 2) = \lambda_0(t)e^{2\beta}$.

If $\beta = 1$, we get that $e^{\beta} = e^1 = 2.73$. This means that an increase of the TNM stage by one level increases the hazard 2.73 times.

We note that the hazard functions for TNM II and IIIa are specified relative to the hazard function of TNM I.

For another parametrization, $X = 2, 1, 0$ for TNM= I, II, IIIa.

We get

- hazard function for TNM IIIa: $\lambda(t|X = 0) = \lambda_0(t)$.
- hazard function for TNM II: $\lambda(t|X = 1) = \lambda_0(t)e^{\beta}$.
- hazard function for TNM I: $\lambda(t|X = 2) = \lambda_0(t)e^{2\beta}$.

The hazard functions for TNM I and II are taken relative to the hazard of TNM IIIa.

The reference level $(X = 0)$ has changed!

Estimation of β

In the Cox's model,

$$\lambda(t|\mathbf{X}) = \lambda_0(t)e^{\beta^t \mathbf{X}},$$

we treat the unknown baseline hazard $\lambda_0(t)$ as a nuisance parameter.

Therefore we need an estimation method for the parameters β without estimating $\lambda_0(t)$.

Hence, we will construct a partial likelihood.

Our data consists a sample of triplets $(T_i, \delta_i, \mathbf{X}_i)$, $i = 1, \ldots, n$ where $\mathbf{X}_i$ is a vector which contains the values of $\mathbf{X}$ for individual i.

We assume

- Given $\mathbf{X}_i$, the lifetime and the censoring time are independent (non-informative censoring).
- Let $\tau_1 < \tau_2 < \ldots < \tau_D$ be the D ordered distinct death times.
- We assume that there are **no tied** death times.

Let us define by

- I_j the identity of the individual who failed at time τ_j.
- V_j the time of the jth failure (τ_j) and all information about censoring in $[\tau_{j-1}, \tau_j[$.

The observable data $(T_i, \delta_i, \mathbf{X}_i)$ is represented by $\{I_j\}$ and $\{V_j\}$. Hence

$$\begin{aligned} P(\text{Data}) &= P(\{I_1, V_1, \ldots, I_D, V_D\}) \\ &= P(\{I_1, V_1\}) \times P(\{I_2, V_2\}|\{I_1, V_1\}) \times \ldots \\ &\quad \times P(\{I_D, V_D\}|\{I_1, V_1, \ldots, I_{D-1}, V_{D-1}\}) \\ &= \prod_{j=1}^{D} P(I_j|\{I_1, V_1, \ldots, I_{j-1}, V_{j-1}, V_j\}) \\ &\quad \times P(V_j|\{I_1, V_1, \ldots, I_{j-1}, V_{j-1}\}) \end{aligned}$$

Due to the non-informative censoring, the second term does not add much information about the parameters β.

Hence, we define the partial likelihood as

$$\begin{aligned} L^{\text{partial}}(\beta) &= \prod_{j=1}^{D} P(I_j|\{I_1, V_1, \ldots, I_{j-1}, V_{j-1}, V_j\}) \\ &= \prod_{j=1}^{D} P(I_j|H_j) \end{aligned}$$

where H_j is the "history" of the data, up to jth failure and including the failure time, but not the identity of the failing.

At each failure, we note that the quantity $P(I_j|H_j)$ is the conditional probability that a specific individual fails at time τ_j, given all the individuals that had not fail before τ_j.

We denote by $\mathcal{R}(t)$ the set of all the individuals under study just prior to time t.

$$
\begin{aligned}
P(I_j|H_j) &= P(\text{individual } I_j \text{ fails}|\text{one individual fails in } \mathcal{R}(\tau_j)) \\
&= \frac{P(\text{individual } j \text{ fails}|\text{at risk at } \tau_j)}{\sum_{l\in\mathcal{R}(\tau_j)} P(\text{individual } l \text{ fails}|\text{at risk at } \tau_j)} \\
&= \frac{\lambda(\tau_j|\mathbf{X}_j)d\tau_j}{\sum_{l\in\mathcal{R}(\tau_j)} \lambda(\tau_j|\mathbf{X}_l)d\tau_j} = \frac{\lambda_0(\tau_j)e^{\beta^t\mathbf{X}_j}}{\sum_{l\in\mathcal{R}(\tau_j)} \lambda_0(\tau_j)e^{\beta^t\mathbf{X}_l}} \\
&= \frac{e^{\beta^t\mathbf{X}_j}}{\sum_{l\in\mathcal{R}(\tau_j)} e^{\beta^t\mathbf{X}_l}}
\end{aligned}
$$

We get as partial likelihood,

$$L^{\text{partial}}(\beta) = \prod_{j=1}^{D} \frac{e^{\beta^t \mathbf{X}_j}}{\sum_{l \in \mathcal{R}(\tau_j)} e^{\beta^t \mathbf{X}_l}}.$$

Some remarks:

- Contributions only at uncensored times.

$$L^{\text{partial}}(\beta) = \prod_{j=1}^{n} \left[\frac{e^{\beta^t \mathbf{X}_j}}{\sum_{l \in \mathcal{R}(\tau_j)} e^{\beta^t \mathbf{X}_l}} \right]^{\delta_j}.$$

- The partial likelihood is **not** a product of independent terms, but of conditional probabilities. Hence, Partial likelihood $\neq$ Usual likelihood.

Although not a usual likelihood, the regular likelihood properties are still valid!

So we have that,

- Log-partial likelihood

$$l(\beta) = \sum_{j=1}^{D} \left[\beta^t \mathbf{X}_j - \log \left[\sum_{l \in \mathcal{R}(\tau_j)} e^{\beta^t \mathbf{X}_l} \right] \right].$$

- Partial likelihood score equation

$$U_i(\beta) = \frac{\partial}{\partial \beta_i} l(\beta) = \sum_{j=1}^{D} \left[\mathbf{X}_{ji} - \frac{\sum_{l \in \mathcal{R}(\tau_j)} \mathbf{X}_{li} e^{\beta^t \mathbf{X}_l}}{\sum_{l \in \mathcal{R}(\tau_j)} e^{\beta^t \mathbf{X}_l}} \right].$$

- Information matrix

$$I_{ik} = -\frac{\partial^2}{\partial\beta_i \partial\beta_k} l(\beta) =$$

$$\sum_{j=1}^{D} \left[\frac{\sum\limits_{l\in\mathcal{R}(\tau_j)} \mathbf{X}_{li}\mathbf{X}_{lk} e^{\beta^t \mathbf{X}_l}}{\sum\limits_{l\in\mathcal{R}(\tau_j)} e^{\beta^t \mathbf{X}_l}} - \frac{\sum\limits_{l\in\mathcal{R}(\tau_j)} \mathbf{X}_{li} e^{\beta^t \mathbf{X}_l} \sum\limits_{l\in\mathcal{R}(\tau_j)} \mathbf{X}_{lk} e^{\beta^t \mathbf{X}_l}}{\left(\sum\limits_{l\in\mathcal{R}(\tau_j)} e^{\beta^t \mathbf{X}_l} \right)^2} \right]$$

Furthermore, we note that

$$\hat{\beta} - \beta \sim N(0, I^{-1}).$$

Based on these quantities, we derive three main test for the global hypothesis $H_0 : \beta = \beta_0$. Under H_0, we get

- Wald-test

$$X^2_W = (b - \beta_0)^t I(b)(b - \beta_0) \sim \chi^2_p$$

- Score test

$$X^2_S = (U_1(\beta_0), \ldots, U_p(\beta_0))^t I^{-1}(\beta_0)(U_1(\beta_0), \ldots, U_p(\beta_0)) \sim \chi^2_p$$

- (Partial) Likelihood ratio

$$X^2_{LR} = 2l(b) - 2l(\beta_0) \sim \chi^2_p$$

Using SAS

```
data nsclc;
input Survtime Survind tnm expres;
cards;
24.51000023      1      2      1
27.12999916      1      2      1
...
;
run;

proc phreg data=nsclc;
model Survtime*Survind(0)=expres;
run;
```

The PHREG Procedure

Model Information

Data Set	WORK.NSCLC
Dependent Variable	Survtime
Censoring Variable	Survind
Censoring Value(s)	0
Ties Handling	BRESLOW

Number of Observations Read	102
Number of Observations Used	102

Summary of the Number of Event and Censored Values

Total	Event	Censored	Percent Censored
102	49	53	51.96

Convergence Status

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

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	400.456	393.305
AIC	400.456	395.305
SBC	400.456	397.197

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	7.1509	1	0.0075
Score	7.1051	1	0.0077
Wald	6.7640	1	0.0093

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
expres	1	0.78590	0.30218	6.7640	0.0093	2.194

Conclusion: expression of P53 increases mortality hazard approximately 2 times as compared to no expression.

Using R

```
> fit<-coxph(Surv(survtime,survind)~expres,data=NSCLC)
> summary(fit)
Call: coxph(formula = Surv(survtime, survind) ~ expres, data = NSCLC)

n= 102
        coef exp(coef) se(coef)   z      p
expres 0.786      2.19    0.302 2.6 0.0093

       exp(coef) exp(-coef) lower .95 upper .95
expres      2.19      0.456      1.21      3.97

Rsquare= 0.068   (max possible= 0.98 )
Likelihood ratio test= 7.15  on 1 df,   p=0.0075
Wald test            = 6.76  on 1 df,   p=0.0093
Score (logrank) test = 7.11  on 1 df,   p=0.00769
```

Using the normality of β, we have a 95% C.I for β:

$$[b - 1.96s.e., b + 1.96s.e.].$$

Hence, the 95% C.I. for the hazard ratio is:

$$[e^{b-1.96s.e.}, e^{b+1.96s.e.}].$$

Adjustments for ties

Sofar, we assumed no tied death times.

In case of ties, we have to adjust the partial likelihood.
Therefore we define

- $\tau_1 < \tau_2 < \ldots < \tau_D$ as the D ordered distinct death times.
- d_j the number of failures at τ_j.
- $I_{j_1}, \ldots, I_{j_d}$ the identities of the individuals who failed at time τ_j.
- H_j is the "history" of the data, up to jth failure and including the failure time, but not the identities of the failing.

The partial likelihood is now defined as

$$L^{\text{partial}}(\beta) = \prod_{j=1}^{D} P(I_{j_1}, \ldots, I_{j_d} | H_j).$$

If the lifetimes are continuous, we find the exact likelihood,

$$L_1 = \prod_{j=1}^{D} \left\{ \int_0^{\infty} \prod_{k \in D_j} \left[1 - \exp\left(\frac{e^{\beta^t X_k}}{\sum_{l \in \mathcal{R}^*(\tau_j)} e^{\beta^t X_l}} t \right) \right] \exp(-t) dt \right\}$$

where $\mathcal{R}^*(\tau_j)$ denote the set of individuals whose event or censored times exceed τ_j or whose censored times are equal to τ_j and D_j is the set of individuals that fail at τ_j.

If the number of ties increases, the denominator is impossible to calculate. Therefore several authors considered approximations

- Breslow

$$L_2(\beta) = \prod_{j=1}^{D} \frac{e^{\beta^t \sum_{k \in D_j} \mathbf{X}_k}}{\left(\sum_{l \in \mathcal{R}(\tau_j)} e^{\beta^t \mathbf{X}_l} \right)^{d_j}}.$$

- Efron

$$L_3(\beta) = \prod_{j=1}^{D} \frac{e^{\beta^t \sum_{k \in D_j} \mathbf{X}_k}}{\prod_{k=1}^{d_j} \left(\sum_{l \in \mathcal{R}(\tau_j)} e^{\beta^t \mathbf{X}_l} - \frac{k-1}{d_j} \sum_{l \in D_j} e^{\beta^t \mathbf{X}_l} \right)}.$$

If the lifetimes are discrete, we get the discrete partial likelihood,

$$L_4(\beta) = \prod_{j=1}^{D} \frac{e^{\beta^t \sum_{k \in D_j} \mathbf{X}_k}}{\sum_{q \in \mathcal{Q}(\tau_j)} e^{\beta^t \sum_{l=1}^{d_j} \mathbf{X}_{ql}}}$$

where $\mathcal{Q}(\tau_j)$ denote the set of all subsets of d_j individuals selected from the risk set $\mathcal{R}(\tau_j)$.

We note that each of these partial likelihoods reduces to the original partial likelihood when there are no ties.

Example: Performance testing

- Test-subject were asked to perform a certain test and the time needed was recorded.
- 3 different noise distractions are applied.
- Did the decreasing noise distractions influence the hazard of the time to finish the test?

Noise level		
9.0	10.0	12.0
9.5	12.0	12.0^+
9.0	12.0^+	12.0^+
8.5	11.0	12.0^+
10.0	12.0	12.0^+
10.5	10.5	12.0^+

In SAS, the Breslow method is the default.

```
data noise;
input Time censor level;
cards;
9.0 1 1
9.5 1 1
9.0 1 1
...
12.0 0 3
;
run;

proc phreg data=noise;
model Time*censor(0)=level;
run;

proc phreg data=noise;
model Time*censor(0)=level/ties=discrete;
run;

proc phreg data=noise;
model Time*censor(0)=level/ties=efron;
run;

proc phreg data=noise;
model Time*censor(0)=level/ties=exact;
run;
```

Ties Handling BRESLOW

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
level	1	-2.10429	0.62384	11.3781	0.0007	0.122

Ties Handling DISCRETE

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
level	1	-2.66767	0.80298	11.0370	0.0009	0.069

Ties Handling EFRON

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
level	1	-2.24966	0.63592	12.5149	0.0004	0.105

Ties Handling EXACT

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
level	1	-2.45815	0.73994	11.0364	0.0009	0.086

In R, the Efron method is the default.

```
> summary(coxph(Surv(Time,Censor)~Level))
Call: coxph(formula = Surv(Time, Censor) ~ Level)

n= 18
       coef exp(coef) se(coef)     z     p
Level -2.25     0.105    0.636 -3.54 4e-04

      exp(coef) exp(-coef) lower .95 upper .95
Level     0.105       9.48    0.0303     0.367


> summary(coxph(Surv(Time,Censor)~Level,method="breslow"))
Call: coxph(formula = Surv(Time, Censor) ~ Level, method = "breslow")

n= 18
       coef exp(coef) se(coef)     z       p
Level -2.10     0.122    0.624 -3.37 0.00074

      exp(coef) exp(-coef) lower .95 upper .95
Level     0.122        8.2    0.0359     0.414


> summary(coxph(Surv(Time,Censor)~Level,method="exact"))
Call: coxph(formula = Surv(Time, Censor) ~ Level, method = "exact")

n= 18
       coef exp(coef) se(coef)     z       p
Level -2.67    0.0694    0.803 -3.32 0.00089

      exp(coef) exp(-coef) lower .95 upper .95
Level    0.0694       14.4    0.0144     0.335
```

Example: NSCLC

- Laudanski et al., Eur Respir J (2001).
- In this study, we had 102 patients who were operated from long cancer.
- The severity of the cancer was expressed in three TNM (Tumor, Nodes, Metastasis) categories: I, II, IIIa.
- The expression of the P53 protein was found from tumor biopsies.

Sofar, expression of P53 has a positive influence on the hazard, however we did not take the confounder TNM into account.

Model 1: $\lambda(t|\mathbf{X}) = \lambda_0(t)e^{\beta_1 \times \mathbf{expres} + \beta_2 \times \mathbf{TNM}}$

```
> fit<-coxph(Surv(survtime,survind)~expres+tnm,data=NSCLC)
> summary(fit)
Call: coxph(formula = Surv(survtime, survind) ~ expres + tnm, data = NSCLC)

n= 102, number of events= 49

         coef exp(coef) se(coef)     z Pr(>|z|)
expres 0.7345    2.0845   0.3032 2.423   0.0154 *
tnm    1.0725    2.9227   0.2546 4.213 2.52e-05 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

       exp(coef) exp(-coef) lower .95 upper .95
expres     2.085     0.4797     1.151     3.776
tnm        2.923     0.3421     1.775     4.814

Concordance= 0.7  (se = 0.044 )
Rsquare= 0.259   (max possible= 0.98 )
Likelihood ratio test= 30.58  on 2 df,   p=2.286e-07
Wald test            = 23.75  on 2 df,   p=6.947e-06
Score (logrank) test = 26.95  on 2 df,   p=1.405e-06
```

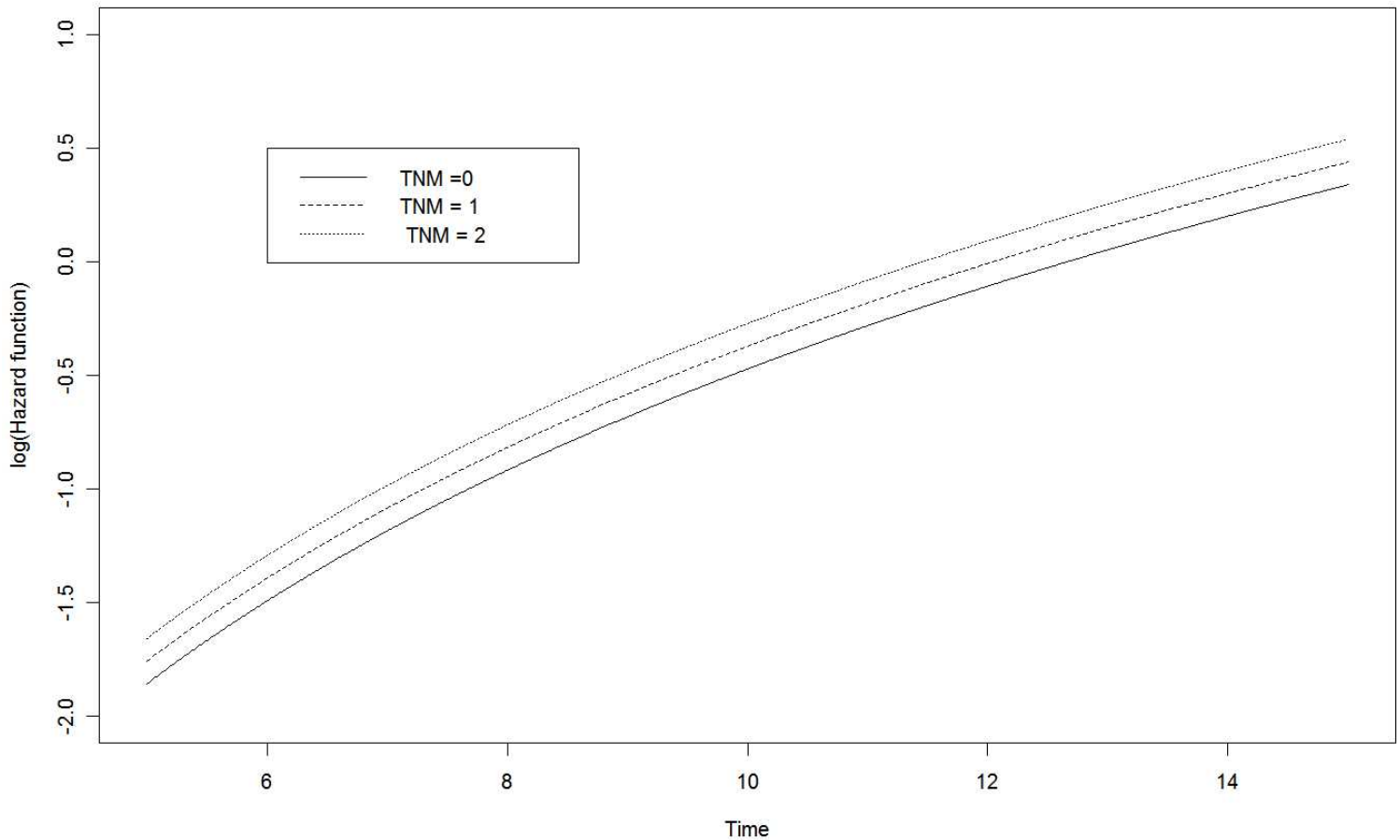
```
proc phreg data=nsclc;
model Survtime*Survind(0)=expres TNM;
run;
```

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
expres	1	0.73452	0.30316	5.8702	0.0154	2.084
tnm	1	1.07226	0.25455	17.7438	<.0001	2.922

Conclusion: Influence of expression P53 changes slightly, TNM has a large positive influence.

However, for different tumor-categories, the change in expression of P53 remains the same!

$$\frac{\lambda(t|\text{expres}=1,\text{TNM}=1)}{\lambda(t|\text{expres}=0,\text{TNM}=1)} = \frac{\lambda(t|\text{expres}=1,\text{TNM}=2)}{\lambda(t|\text{expres}=0,\text{TNM}=2)} = e^{\beta_1}.$$


TNM =0
TNM = 1
TNM = 2
log(Hazard function)
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0
6
8
10
12
14
Time

Model 2: $\lambda(t|\mathbf{X}) = \lambda_0(t)e^{\beta_1\mathbf{expres}+\beta_2\mathbf{TNM}+\beta_3\mathbf{expres}*\mathbf{TNM}}$

```
> fit<-coxph(Surv(survtime,survind)~expres*tnm,data=NSCLC)
> summary(fit)
Call: coxph(formula = Surv(survtime, survind) ~ expres * tnm, data = NSCLC)

n= 102, number of events= 49

              coef exp(coef) se(coef)      z Pr(>|z|)
expres      3.4814   32.5052   1.7444  1.996  0.04595 *
tnm         1.7948    6.0180   0.5612  3.198  0.00138 **
expres:tnm -1.0186    0.3611   0.6198 -1.643  0.10029
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

           exp(coef) exp(-coef) lower .95 upper .95
expres       32.5052    0.03076    1.0645   992.541
tnm           6.0180    0.16617    2.0034    18.078
expres:tnm    0.3611    2.76925    0.1072     1.217

Concordance= 0.709  (se = 0.044 )
Rsquare= 0.283   (max possible= 0.98 )
Likelihood ratio test= 33.87  on 3 df,   p=2.11e-07
Wald test            = 18.71  on 3 df,   p=0.0003131
Score (logrank) test = 27.02  on 3 df,   p=5.827e-06
```

```
data nsclc1;
set nsclc;
exTNM=expres*TNM;
run;

proc phreg data=nsclc1;
model Survtime*Survind(0)=expres TNM exTNM;
run;
```

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
expres	1	3.48139	1.74435	3.9833	0.0460	32.505
tnm	1	1.79476	0.56120	10.2276	0.0014	6.018
exTNM	1	-1.01857	0.61978	2.7009	0.1003	0.361

Conclusion: Influence of expression P53 changes drastically by interaction with TNM. However, interaction is not significant.

Model 3: $\lambda(t|\mathbf{X}) = \lambda_0(t)e^{\beta_1\mathbf{expres}+\beta_2(\mathbf{TNM}=1)+\beta_3(\mathbf{TNM}=2)}$

```
> fit<-coxph(Surv(survtime,survind)~expres+(tnm==1)+(tnm==2),data=NSCLC)
> summary(fit)
Call: coxph(formula = Surv(survtime, survind) ~ expres + (tnm == 1) + (tnm == 2), data = NSCLC)

n= 102, number of events= 49

                coef exp(coef) se(coef)      z Pr(>|z|)
expres        0.7594    2.1369   0.3050  2.490 0.012790 *
tnm == 1TRUE -1.7225    0.1786   0.4995 -3.449 0.000563 ***
tnm == 2TRUE -1.5788    0.2062   0.4251 -3.714 0.000204 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

             exp(coef) exp(-coef) lower .95 upper .95
expres          2.1369      0.468   1.17532    3.8852
tnm == 1TRUE    0.1786      5.598   0.06711    0.4754
tnm == 2TRUE    0.2062      4.849   0.08963    0.4744

Concordance= 0.709  (se = 0.044 )
Rsquare= 0.279   (max possible= 0.98 )
Likelihood ratio test= 33.31  on 3 df,   p=2.768e-07
Wald test            = 27.71  on 3 df,   p=4.173e-06
Score (logrank) test = 32.63  on 3 df,   p=3.849e-07
```

```
> fit<-coxph(Surv(survtime,survind)~expres,data=NSCLC)
> summary(fit)
Call: coxph(formula = Surv(survtime, survind) ~ expres, data = NSCLC)

n= 102, number of events= 49

         coef exp(coef) se(coef)     z Pr(>|z|)
expres 0.7859    2.1945   0.3022 2.601   0.0093 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

       exp(coef) exp(-coef) lower .95 upper .95
expres     2.194     0.4557     1.214     3.968

Concordance= 0.599  (se = 0.039 )
Rsquare= 0.068   (max possible= 0.98 )
Likelihood ratio test= 7.15  on 1 df,   p=0.007493
Wald test            = 6.76  on 1 df,   p=0.009298
Score (logrank) test = 7.11  on 1 df,   p=0.007687

> fit1<-coxph(Surv(survtime,survind)~expres+(tnm==1)+(tnm==2),data=NSCLC)

> anova(fit1,fit)
Analysis of Deviance Table
Cox model: response is  Surv(survtime, survind)
Model 1: ~ expres + (tnm == 1) + (tnm == 2)
Model 2: ~ expres
   loglik  Chisq Df P(>|Chi|)
1 -183.57
2 -196.65 26.161  2 2.086e-06 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

```
data nsclc1;
set nsclc;
TNM1=0;
TNM2=0;
If TNM=1 then TNM1=1;
If TNM=2 then TNM2=1;
run;

proc phreg data=nsclc1;
model Survtime*Survind(0)=expres TNM1 TNM2;
TNM: test TNM1=0,TNM2=0;
TNMC: test TNM2=2*TNM1;
run;
```

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
expres	1	0.75936	0.30502	6.1980	0.0128	2.137
TNM1	1	-1.72246	0.49947	11.8927	0.0006	0.179
TNM2	1	-1.57884	0.42512	13.7931	0.0002	0.206

Linear Hypotheses Testing Results

Label	Wald Chi-Square	DF	Pr > ChiSq
TNM	21.4783	2	<.0001
TNMC	3.4425	1	0.0635

Conclusion: Considering TNM as a class variable, we still see its significant influence. A local test shows that we can take it as a continuous variable.

Still, we assume that TNM has a proportional influence on the hazard function.

- For some covariates, the proportional hazard assumption can be fulfilled, but not for others.
- In such a situation, one can use the Cox model, using modified baseline hazards.

Assume that the hazard is proportional for a covariate Y, but for X.

We take a stratified Cox's model,

$$\lambda(t|X,Y) = \lambda_{X=x,0}(t)e^{\beta Y}.$$

Note that the baseline hazard functions are different for strata defined by the levels of X, but the effect of Y is still expressed as the proportional change of the hazard function (for a fixed level of X).

Model 4: $\lambda(t|\mathbf{X}) = \lambda_{\mathbf{TNM},0}(t)e^{\beta \times \mathbf{expres}}$

```
> fit<-coxph(Surv(survtime,survind)~expres+strata(tnm),data=NSCLC)
> summary(fit)
Call: coxph(formula = Surv(survtime, survind) ~ expres + strata(tnm), data = NSCLC)

n= 102, number of events= 49

         coef exp(coef) se(coef)     z Pr(>|z|)
expres 0.7404    2.0968   0.3076 2.407   0.0161 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

       exp(coef) exp(-coef) lower .95 upper .95
expres     2.097     0.4769     1.147     3.832

Concordance= 0.597  (se = 0.058 )
Rsquare= 0.059   (max possible= 0.949 )
Likelihood ratio test= 6.16  on 1 df,   p=0.0131
Wald test            = 5.79  on 1 df,   p=0.01609
Score (logrank) test = 6.05  on 1 df,   p=0.01387
```

```
proc phreg data=nsclc1;
model Survtime*Survind(0)=expres;
strata TNM;
run;
```

Summary of the Number of Event and Censored Values

Stratum	tnm	Total	Event	Censored	Percent Censored
1	1	21	5	16	76.19
2	2	27	7	20	74.07
3	3	54	37	17	31.48
Total		102	49	53	51.96

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
expres	1	0.74029	0.30763	5.7910	0.0161	2.097

Next to the estimation of the parameters β, we want to estimate the survival function at specific covariate values.

Hereto we consider two methods:

- Kalbfleisch-Prentice method (no ties):

$$\hat{S}(t|\mathbf{X}) = S_0(t)^{\exp(\hat{\beta}^t\mathbf{X})},\ S_0(t) = \prod_{t_i \leq t}\left[1 - \frac{\exp(\hat{\beta}^t\mathbf{X}_i)}{\sum_{l\in\mathcal{R}(t_i)}\exp(\hat{\beta}^t\mathbf{X}_l)}\right]$$

- Breslow method:

$$\hat{S}(t|\mathbf{X}) = S_0(t)^{\exp(\hat{\beta}^t\mathbf{X})},\ S_0(t) = \prod_{t_i \leq t}\exp\left[-\frac{d_i}{\sum_{l\in\mathcal{R}(t_i)}\exp(\hat{\beta}^t\mathbf{X}_l)}\right]$$

```
data teken;
input x;
datalines;
0.00
1.00
;

proc phreg data=test;
model time*status(0)=x/ties=efron;
baseline covariates=teken out=predd survival=_all_/method=pl;
run;

proc print data=predd;
run;

proc phreg data=test;
model time*status(0)=x/ties=efron;
baseline covariates=teken out=predd1 survival=_all_;
run;

proc print data=predd1;
run;
```

Obs	x	time	Survival	StdErr Survival	Lower Survival	Upper Survival
1	0	0	1.00000	.	.	.
2	0	1	0.89962	0.09941	0.72444	1.00000
3	0	2	0.63352	0.18915	0.35287	1.00000
4	0	3	0.45826	0.21153	0.18544	1.00000
5	0	4	0.00000	0.00000	.	.
6	1	0	1.00000	.	.	.
7	1	1	0.84562	0.13059	0.62478	1.00000
8	1	2	0.48502	0.17082	0.24320	0.96726
9	1	3	0.29027	0.14425	0.10960	0.76881
10	1	4	0.00000	0.00000	.	.

Obs	x	time	Survival	StdErr Survival	Lower Survival	Upper Survival
1	0	0	1.00000	.	.	.
2	0	1	0.90720	0.10024	0.73055	1
3	0	2	0.68444	0.20435	0.38123	1
4	0	3	0.54840	0.25313	0.22192	1
5	0	4	0.20175	0.22220	0.02330	1
6	1	0	1.00000	.	.	.
7	1	1	0.85695	0.13234	0.63315	1
8	1	2	0.54825	0.19309	0.27491	1
9	1	3	0.38585	0.19175	0.14568	1
10	1	4	0.07907	0.14905	0.00196	1

```
> test1 <- list(time=c(4,3,1,1,2,2,3),
+               status=c(1,1,1,0,1,1,0),
+               x=c(0,2,1,1,1,0,0))
```

```
> summary(survfit(coxph(Surv(time, status) ~ x, test1),type="kalbfleisch-prentice",
newdata=data.frame(x=c(0,1))))
Call: survfit(formula = coxph(Surv(time, status) ~ x, test1), newdata = data.frame(x = c(0,1)),
 type = "kalbfleisch-prentice")

time n.risk n.event survival1 survival2
1      7       1     0.900     0.846
2      5       2     0.634     0.485
3      3       1     0.458     0.290
4      1       1     0.000     0.000
```

```
> summary(survfit(coxph(Surv(time, status) ~ x, test1),type="breslow",newdata=data.frame(x=c(0,1))))
Call: survfit(formula = coxph(Surv(time, status) ~ x, test1), newdata = data.frame(x = c(0,1)),
 type = "breslow")

time n.risk n.event survival1 survival2
1      7       1     0.907    0.8569
2      5       2     0.684    0.5482
3      3       1     0.548    0.3859
4      1       1     0.202    0.0791
```

Time-dependent covariates

Suppose:

- After a kidney transplantation, we look at the time until the host body rejects the organ.
- Every month, a patient has a check-up and blood pressure, white cell count,... are recorded.
- we are interested how these time-dependent covariates influence the hazard of rejection time.

Hence, we have $(T_i, \delta_i, \{\mathbf{X}_i(t), 0 \leq t \leq T_i\})$, $i = 1, \ldots, n$.

Note that fixed-time covariates are a special case, namely,

$$\mathbf{X}(t) = \mathbf{X}(0), \quad \forall t > 0.$$

Extending the Cox's model to accommodate for time-dependent covariates, we assume that

$$\lambda(t|\mathbf{X}(t)) = \lambda_0(t)e^{\beta^t \mathbf{X}(t)}.$$

To estimate the parameters β, we extend the partial likelihood.

We assume

- The value of $\mathbf{X}_i(t)$ is known for any time at which the subject is at risk.
- Given $\mathbf{X}_i(t)$, the lifetime and the censoring time are independent (non-informative censoring).
- Let $\tau_1 < \tau_2 < \ldots < \tau_D$ be the D ordered distinct death times.
- We assume that there are **no tied** death times.

We get as partial likelihood,

$$L^{\text{partial}}(\beta) = \prod_{j=1}^{D} \frac{e^{\beta^t \mathbf{X}_j(\tau_j)}}{\sum_{l \in \mathcal{R}(\tau_j)} e^{\beta^t \mathbf{X}_l(\tau_j)}}.$$

Some remarks:

- For each individual, we only need the values of $\mathbf{X}(t)$ at the uncensored death times.
- When there are ties, we extend as before one of the previous partial likelihoods.

Example: Cancer in rodents

- Forty-five rodents were randomly assigned to three dose groups of a tumor-promoting agent.
- The rodents were examined every week for the number of papillomas (weeks 27, 34, 37, 41, 43, 45, 46, 47, 49, 50, 51, 53, 65, 67 and 71).
- Researchers are interested in the time until death of cancer and how this was influenced by dose after adjusting for the number of papillomas.

```
> rodent<-read.table("C:/werk/Roel/Onderwijs/Theorie/G0B67AStatAnalReliaSurvData/
Cursus/Rodent.txt",header=T,sep=";")
> rodent[1:5,]
   Id Time Dead Dose P1 P2 P3 P4 P5 P6 P7 P8 P9 P10 P11 P12 P13 P14 P15
 1  1   47    1    1  0  5  6  8 10 10 10 10 NA  NA  NA  NA  NA  NA  NA
 2  2   71    1    1  0  0  0  0  0  0  0  0  1   1   1   1   1   1   1
 3  3   81    0    1  0  1  1  1  1  1  1  1  1   1   1   1   1   1   1
 4  4   81    0    1  0  0  0  0  0  0  0  0  0   0   0   0   0   0   0
 5  5   81    0    1  0  0  0  0  0  0  0  0  0   0   0   0   0   0   0

> n<-dim(rodent)[1]
> Date0<-rep(0,n)
> P0<-rep(0,n)
> Date1<-rep(27,n)
> Date2<-rep(34,n)
> Date3<-rep(37,n)
> Date4<-rep(41,n)
> Date5<-rep(43,n)
> Date6<-rep(45,n)
> Date7<-rep(46,n)
> Date8<-rep(47,n)
> Date9<-rep(49,n)
> Date10<-rep(50,n)
> Date11<-rep(51,n)
> Date12<-rep(53,n)
> Date13<-rep(65,n)
> Date14<-rep(67,n)
> Date15<-rep(71,n)
> rodent1<-data.frame(rodent,P0,Date0,Date1,Date2,Date3,Date4,Date5,Date6,Date7,Date8,Date9,Date10,
Date11,Date12,Date13,Date14,Date15)
> rodentlong<-tmerge(rodent1[,1:4],rodent1,id=Id,status=event(Time,Dead),pap=tdc(Date0,P0),
+ pap=tdc(Date1,P1),pap=tdc(Date2,P2),pap=tdc(Date3,P3),pap=tdc(Date4,P4),pap=tdc(Date5,P5),
+ pap=tdc(Date6,P6),pap=tdc(Date7,P7),pap=tdc(Date8,P8),pap=tdc(Date9,P9),pap=tdc(Date10,P10),
+ pap=tdc(Date11,P11),pap=tdc(Date12,P12),pap=tdc(Date13,P13),pap=tdc(Date14,P14),pap=tdc(Date15,P15))
```

```
> rodentlong[1:20,4:9]
   Dose id tstart tstop status pap
1     1  1      0    27      0   0
2     1  1     27    34      0   0
3     1  1     34    37      0   5
4     1  1     37    41      0   6
5     1  1     41    43      0   8
6     1  1     43    45      0  10
7     1  1     45    46      0  10
8     1  1     46    47      1  10
9     1  2      0    27      0   0
10    1  2     27    34      0   0
11    1  2     34    37      0   0
12    1  2     37    41      0   0
13    1  2     41    43      0   0
14    1  2     43    45      0   0
15    1  2     45    46      0   0
16    1  2     46    47      0   0
17    1  2     47    49      0   0
18    1  2     49    50      0   1
19    1  2     50    51      0   1
20    1  2     51    53      0   1
```

```
> fit<-coxph(Surv(tstart,tstop,status)~Dose+pap,data=rodentlong)
> summary(fit)
Call: coxph(formula = Surv(tstart, tstop, status) ~ Dose + pap, data = rodentlong)

n= 428, number of events= 25

        coef exp(coef) se(coef)     z Pr(>|z|)
Dose 0.09132   1.09562  0.05582 1.636  0.10183
pap  0.10303   1.10852  0.03163 3.257  0.00113 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

     exp(coef) exp(-coef) lower .95 upper .95
Dose     1.096     0.9127    0.9821     1.222
pap      1.109     0.9021    1.0419     1.179

Concordance= 0.809  (se = 0.064 )
Rsquare= 0.045   (max possible= 0.321 )
Likelihood ratio test= 19.75  on 2 df,   p=5.139e-05
Wald test            = 18.27  on 2 df,   p=0.0001078
Score (logrank) test = 23.57  on 2 df,   p=7.616e-06
```

```
option nocenter;
data rodent;
infile datalines missover;
input ID Time Dead Dose P1-P15;
label ID='Subject ID'; datalines;
1 47 1  1.0  0  5  6  8 10 10 10 10
2 71 1  1.0  0  0  0  0  0  0  0  0  1  1  1  1 1 1 1
3 81 0  1.0  0  1  1  1  1  1  1  1  1  1  1  1 1 1 1
...
44 37 1 10.0  0  1  1
45 43 1 10.0  9 19 19 19 19
;

proc phreg data=rodent;
model Time*Dead(0)=Dose NPap;
array pp{*} P1-P14;
array tt{*} t1-t15;
t1 = 27;
t2 = 34;
t3 = 37;
t4 = 41;
t5 = 43;
t6 = 45;
t7 = 46;
t8 = 47;
t9 = 49;
t10= 50;
t11= 51;
t12= 53;
t13= 65;
t14= 67;
t15= 71;
if Time <  tt[1]  then NPap=0;
else if time >= tt[15] then NPap=P15;
else do i=1 to dim(pp);
if tt[i] <= Time < tt[i+1] then NPap= pp[i];
end;
run;
```

The PHREG Procedure

Model Information

Data Set	WORK.RODENT
Dependent Variable	Time
Censoring Variable	Dead
Censoring Value(s)	0
Ties Handling	BRESLOW

Number of Observations Read	45
Number of Observations Used	45

Summary of the Number of Event and Censored Values

Total	Event	Censored	Percent Censored
45	25	20	44.44

Convergence Status

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

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	166.793	143.269
AIC	166.793	147.269
SBC	166.793	149.707

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	23.5243	2	<.0001
Score	28.0498	2	<.0001
Wald	21.1646	2	<.0001

Analysis of Maximum Likelihood Estimates

Parameter Variable	DF	Standard Estimate	Error	Chi-Square	Pr > ChiSq	Hazard Ratio
Dose	1	0.06885	0.05620	1.5010	0.2205	1.071
NPap	1	0.11714	0.02998	15.2705	<.0001	1.124

Model building

In the previous examples, we looked at how maximal two covariates influenced the survival time in the Cox's regression model.

However in studies with several covariates, we need a procedure to identify interesting covariates.

In nested models, we used the likelihood ratio test to check whether a covariate was significant or not. Here, we extend this idea and consider the *AIC*-criterium,

$$AIC = -2 \log \hat{L} + \alpha q,$$

in which q is the number of unknown β parameters.

In practice, model building is a combination of

- knowledge of the science
- trial and error, common sense
- automatic variable selection
  - stepwise: forward, backward, both
  - best subsets

A general strategy that is commonly used is,

1. Perform univariate analysis to "screen" potentially significant variables.
2. Fit a multiple model and discard variables that are non-significant.
3. Check whether variables that were dropped before, interactions or higher order terms should be added.

Example: Leukemia

- Bone marrow transplants are a standard treatment for acute leukemia.
- The researchers investigated the time until relapse of leukemia.
- Several covariates: Disease Group, Patient Age (Years), Donor Age (Years), Patient Sex, Donor Sex, Patient CMV Status, Donor CMV Status, Waiting Time to Transplant In Days, FAB grade, Hospital, MTX Use.

```
proc phreg data=bmt1;
model T2*delta2(0)=g1 g2 Z1 Z2 Z3 Z4 Z5 Z6 Z7 Z8 Z91 z92 z93 Z10/selection=forward details sle=0.1
include=2;
run;

proc phreg data=bmt1;
model T2*delta2(0)=g1 g2 Z1 Z2 Z3 Z4 Z5 Z6 Z7 Z8 Z91 z92 z93 Z10/selection=backwards details sls=0.1
include=2;
run;

proc phreg data=bmt1;
model T2*delta2(0)=g1 g2 Z1 Z2 Z3 Z4 Z5 Z6 Z7 Z8 Z91 z92 z93 Z10/selection=stepwise details sle=0.1
sls=0.07 include=2;
run;

proc phreg data=bmt1;
model T2*delta2(0)=g1 g2 Z1 Z2 Z3 Z4 Z5 Z6 Z7 Z8 Z91 z92 z93 Z10/selection=score best=4 include=2;
run;
```

The PHREG Procedure

Model Information

Data Set	WORK.BMT1
Dependent Variable	T2
Censoring Variable	delta2
Censoring Value(s)	0
Ties Handling	BRESLOW

Number of Observations Read	137
Number of Observations Used	137

Summary of the Number of Event and Censored Values

Total	Event	Censored	Percent Censored
137	42	95	69.34

The following variable(s) will be included in each model:

g1 g2

Convergence Status

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

Analysis of Variables Not in the Model

Variable	Score Chi-Square	Pr > ChiSq
Z1	0.0099	0.9209
Z2	0.0580	0.8097
Z3	2.0354	0.1537
Z4	0.8051	0.3696
Z5	0.8466	0.3575
Z6	0.0231	0.8793
Z7	2.6081	0.1063
Z8	11.2959	0.0008
z91	0.2108	0.6462
z92	0.1297	0.7187
z93	0.5913	0.4419
Z10	0.8084	0.3686

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
19.1425	11	0.0586

Step 1. Variable Z8 is entered. The model contains the following explanatory variables:

g1 g2 Z8

Convergence Status

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

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	380.487	353.210
AIC	380.487	359.210
SBC	380.487	364.423

The PHREG Procedure

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	27.2772	3	<.0001
Score	27.8146	3	<.0001
Wald	24.4659	3	<.0001

Analysis of Maximum Likelihood Estimates

Parameter Variable	DF	Standard Estimate	Error	Chi-Square	Pr > ChiSq	Hazard Ratio
g1	1	-1.53927	0.51812	8.8260	0.0030	0.215
g2	1	-0.22154	0.48465	0.2090	0.6476	0.801
Z8	1	1.31812	0.41855	9.9179	0.0016	3.736

Analysis of Variables Not in the Model

Variable	Score Chi-Square	Pr > ChiSq
Z1	0.6505	0.4199
Z2	0.0024	0.9613
Z3	0.7547	0.3850
Z4	0.9355	0.3334
Z5	0.4415	0.5064
Z6	0.0071	0.9329
Z7	2.5716	0.1088
z91	0.0330	0.8558
z92	0.0003	0.9854
z93	1.1483	0.2839
Z10	0.7865	0.3751

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
7.7833	10	0.6500

NOTE: No (additional) variables met the 0.1 level for entry into the model.

Summary of Forward Selection

Step	Variable Entered	Number In	Score Chi-Square	Pr > ChiSq
1	Z8	3	11.2959	0.0008

Model Information

Data Set	WORK.BMT1
Dependent Variable	T2
Censoring Variable	delta2
Censoring Value(s)	0
Ties Handling	BRESLOW

Number of Observations Read	137
Number of Observations Used	137

Summary of the Number of Event and Censored Values

Total	Event	Censored	Percent Censored
137	42	95	69.34

The following variable(s) will be included in each model:

g1 g2

Step 0. The model contains the following variables:

g1 g2 Z1 Z2 Z3 Z4 Z5 Z6 Z7 Z8 z91 z92 z93 Z10

Convergence Status

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

Analysis of Maximum Likelihood Estimates

Parameter Variable	DF	Standard Estimate	Error	Chi-Square	Pr > ChiSq	Hazard Ratio
g1	1	-1.97737	0.60510	10.6789	0.0011	0.138
g2	1	-0.49615	0.53394	0.8635	0.3528	0.609
Z1	1	0.04161	0.03170	1.7223	0.1894	1.042
Z2	1	-0.03104	0.02816	1.2154	0.2703	0.969
Z3	1	-0.30270	0.35289	0.7358	0.3910	0.739
Z4	1	0.39950	0.37151	1.1564	0.2822	1.491
Z5	1	0.14540	0.38000	0.1464	0.7020	1.157
Z6	1	-0.07759	0.36061	0.0463	0.8296	0.925
Z7	1	-0.0009995	0.0007768	1.6555	0.1982	0.999
Z8	1	1.44166	0.44917	10.3015	0.0013	4.228
z91	1	0.63995	0.57090	1.2565	0.2623	1.896
z92	1	0.51117	0.76634	0.4449	0.5048	1.667
z93	1	0.86387	0.59507	2.1075	0.1466	2.372
Z10	0	0	.	.	.	.

Step 1. Variable Z10 is removed because of its redundancy.

Step 2. Variable Z6 is removed. The model contains the following explanatory variables:

g1 g2 Z1 Z2 Z3 Z4 Z5 Z7 Z8 z91 z92 z93

Convergence Status

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

Analysis of Maximum Likelihood Estimates

Parameter Variable	DF	Standard Estimate	Error	Chi-Square	Pr > ChiSq	Hazard Ratio
g1	1	-1.96479	0.60164	10.6650	0.0011	0.140
g2	1	-0.48874	0.53240	0.8427	0.3586	0.613
Z1	1	0.04197	0.03163	1.7606	0.1845	1.043
Z2	1	-0.03203	0.02772	1.3350	0.2479	0.968
Z3	1	-0.28731	0.34564	0.6909	0.4058	0.750
Z4	1	0.40862	0.36893	1.2267	0.2680	1.505
Z5	1	0.11851	0.35856	0.1092	0.7410	1.126
Z7	1	-0.0009920	0.0007736	1.6442	0.1998	0.999
Z8	1	1.45514	0.44548	10.6698	0.0011	4.285
z91	1	0.64716	0.56998	1.2892	0.2562	1.910
z92	1	0.51320	0.76571	0.4492	0.5027	1.671
z93	1	0.86823	0.59393	2.1370	0.1438	2.383

...

Step 11. Variable Z7 is removed. The model contains the following explanatory variables:

g1 g2 Z8

Convergence Status

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

Analysis of Maximum Likelihood Estimates

Parameter Variable	DF	Standard Estimate	Error	Chi-Square	Pr > ChiSq	Hazard Ratio
g1	1	-1.53927	0.51812	8.8260	0.0030	0.215
g2	1	-0.22154	0.48465	0.2090	0.6476	0.801
Z8	1	1.31812	0.41855	9.9179	0.0016	3.736

NOTE: No (additional) variables met the 0.1 level for removal from the model.

Summary of Backward Elimination

Step	Variable Removed	Number In	Wald Chi-Square	Pr > ChiSq
1	Z10	13	.	.
2	Z6	12	0.0463	0.8296
3	Z5	11	0.1092	0.7410
4	z92	10	0.6167	0.4323
5	Z3	9	0.6233	0.4298
6	z91	8	0.5974	0.4396
7	Z4	7	0.7501	0.3864
8	z93	6	1.0015	0.3169
9	Z2	5	1.2315	0.2671
10	Z1	4	0.3947	0.5299
11	Z7	3	2.4189	0.1199

Analysis of Variables Not in the Model

Variable	Score Chi-Square	Pr > ChiSq
Z1	0.0099	0.9209
Z2	0.0580	0.8097
Z3	2.0354	0.1537
Z4	0.8051	0.3696
Z5	0.8466	0.3575
Z6	0.0231	0.8793
Z7	2.6081	0.1063
Z8	11.2959	0.0008
z91	0.2108	0.6462
z92	0.1297	0.7187
z93	0.5913	0.4419
Z10	0.8084	0.3686

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
19.1425	11	0.0586

Step 1. Variable Z8 is entered. The model contains the following explanatory variables:

g1 g2 Z8

Convergence Status

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

Analysis of Variables Not in the Model

Variable	Score Chi-Square	Pr > ChiSq
Z1	0.6505	0.4199
Z2	0.0024	0.9613
Z3	0.7547	0.3850
Z4	0.9355	0.3334
Z5	0.4415	0.5064
Z6	0.0071	0.9329
Z7	2.5716	0.1088
z91	0.0330	0.8558
z92	0.0003	0.9854
z93	1.1483	0.2839
Z10	0.7865	0.3751

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
7.7833	10	0.6500

NOTE: No (additional) variables met the 0.1 level for entry into the model.

Summary of Stepwise Selection

Step	Variable Entered	Number Removed	Score In	Wald Chi-Square	Chi-Square	Pr > ChiSq
1	Z8		3	11.2959	.	0.0008

NOTE: The following variables are not used in the SCORE selection since they are a linear combination of other variables as shown.

Z10 = 1 * z92 + 1 * z93

Regression Models Selected by Score Criterion

Number of Variables	Score Chi-Square	Variables Included in Model
2	16.4771	g1 g2
3	27.8146	g1 g2 Z8
3	19.5130	g1 g2 Z7
3	18.4199	g1 g2 Z3
3	17.3021	g1 g2 Z5
4	30.9681	g1 g2 Z7 Z8
4	29.2600	g1 g2 Z3 Z8
4	28.6084	g1 g2 Z8 z93
4	28.4205	g1 g2 Z4 Z8
5	31.8900	g1 g2 Z3 Z7 Z8
5	31.6531	g1 g2 Z7 Z8 z93
5	31.3819	g1 g2 Z5 Z7 Z8
5	31.3151	g1 g2 Z4 Z7 Z8
6	32.5001	g1 g2 Z3 Z4 Z7 Z8
6	32.4470	g1 g2 Z3 Z7 Z8 z93
6	32.4034	g1 g2 Z3 Z5 Z7 Z8
6	32.0224	g1 g2 Z4 Z7 Z8 z93

```
7     33.0644  g1 g2 Z3 Z4 Z7 Z8 z93
7     32.9565  g1 g2 Z3 Z4 Z5 Z7 Z8
7     32.8969  g1 g2 Z3 Z5 Z7 Z8 z93
7     32.6351  g1 g2 Z3 Z7 Z8 z91 z93

8     33.4602  g1 g2 Z3 Z4 Z5 Z7 Z8 z93
8     33.2658  g1 g2 Z3 Z4 Z7 Z8 z91 z93
8     33.1314  g1 g2 Z2 Z3 Z4 Z7 Z8 z93
8     33.1069  g1 g2 Z3 Z4 Z7 Z8 z92 z93

9     33.6562  g1 g2 Z3 Z4 Z5 Z7 Z8 z91 z93
9     33.5729  g1 g2 Z2 Z3 Z4 Z5 Z7 Z8 z93
9     33.4994  g1 g2 Z3 Z4 Z5 Z6 Z7 Z8 z93
9     33.4965  g1 g2 Z3 Z4 Z7 Z8 z91 z92 z93

10     33.7837  g1 g2 Z3 Z4 Z5 Z7 Z8 z91 z92 z93
10     33.7755  g1 g2 Z1 Z2 Z3 Z4 Z7 Z8 z91 z93
10     33.7660  g1 g2 Z2 Z3 Z4 Z5 Z7 Z8 z91 z93
10     33.7142  g1 g2 Z1 Z2 Z3 Z4 Z5 Z7 Z8 z93

11     34.0330  g1 g2 Z1 Z2 Z3 Z4 Z5 Z7 Z8 z91 z93
11     33.9245  g1 g2 Z1 Z2 Z3 Z4 Z7 Z8 z91 z92 z93
11     33.8610  g1 g2 Z2 Z3 Z4 Z5 Z7 Z8 z91 z92 z93
11     33.8210  g1 g2 Z3 Z4 Z5 Z6 Z7 Z8 z91 z92 z93

12     34.1143  g1 g2 Z1 Z2 Z3 Z4 Z5 Z7 Z8 z91 z92 z93
12     34.0635  g1 g2 Z1 Z2 Z3 Z4 Z5 Z6 Z7 Z8 z91 z93
12     33.9279  g1 g2 Z1 Z2 Z3 Z4 Z6 Z7 Z8 z91 z92 z93
12     33.8841  g1 g2 Z2 Z3 Z4 Z5 Z6 Z7 Z8 z91 z92 z93

13     34.1454  g1 g2 Z1 Z2 Z3 Z4 Z5 Z6 Z7 Z8 z91 z92 z93
```

After fitting a Cox's regression model to a practical data set, it is important to check whether the Cox's regression model is an correct model for this data set.

Like in ordinary regression, we use residuals to assess goodness of fit.

In survival analysis, several types of residuals can be determined:

- Cox-Snel residuals
- Schoenfeld residuals
- Martingale residuals

As a first type of residuals, we consider Cox-Snel residuals.

These residuals are defined as

$$r_i = \hat{H}_0(t_i) \exp(x_i \hat{\beta})$$

where $\hat{H}_0(t_i)$ is the baseline cumulative hazard function at t_i.

If the Cox's regression model is satisfied, we get that r_i is a censored sample of an exponential distribution with lambda 1.

Testing for this, gives a test of model adequacy.

Graphically, we can do this by plotting a cumulative hazard estimate of these residuals. Since $H_0(t) = t$ for an exponential distribution, we should see a straight line.

We concentrate on martingale residuals.

For every individual, we get

$$r_i^m = \delta_i - \hat{H}(T_i), \quad i = 1, \ldots, n$$

where $\hat{H}$ is the fitted cumulative hazard function under the Cox's regression model.

We note that

- The martingale residuals sum to zero.
- In large sample, the martingale residuals are uncorrelated and have an expected value of zero.
- For each individual, the martingale residual looks like the difference of the observed number of deaths in interval $[0, t_i[$ minus expected number under the fitted value.

By plotting these values against index, fitted values or covariates we asses whether there are outliers, the models fits or the functional form of the covariates is satisfied.

A third type of residuals are the Schoenfeld residuals.

From the partial likelihood, we know that the parameter β are estimated from

$$\sum_{i=1}^{d}(x_i - E[x_i|\mathcal{R}(t_i)]) = 0$$

with

$$E[x_i|\mathcal{R}(t_i)] = \frac{\sum_{l\in\mathcal{R}(t_i)} x_l \exp(x_l^t\beta)}{\sum_{l\in\mathcal{R}(t_i)} \exp(x_l^t\beta)}.$$

The Schoenfeld residual are defined as

$$r_i^s = x_i - E[x_i|\mathcal{R}(t_i)]$$

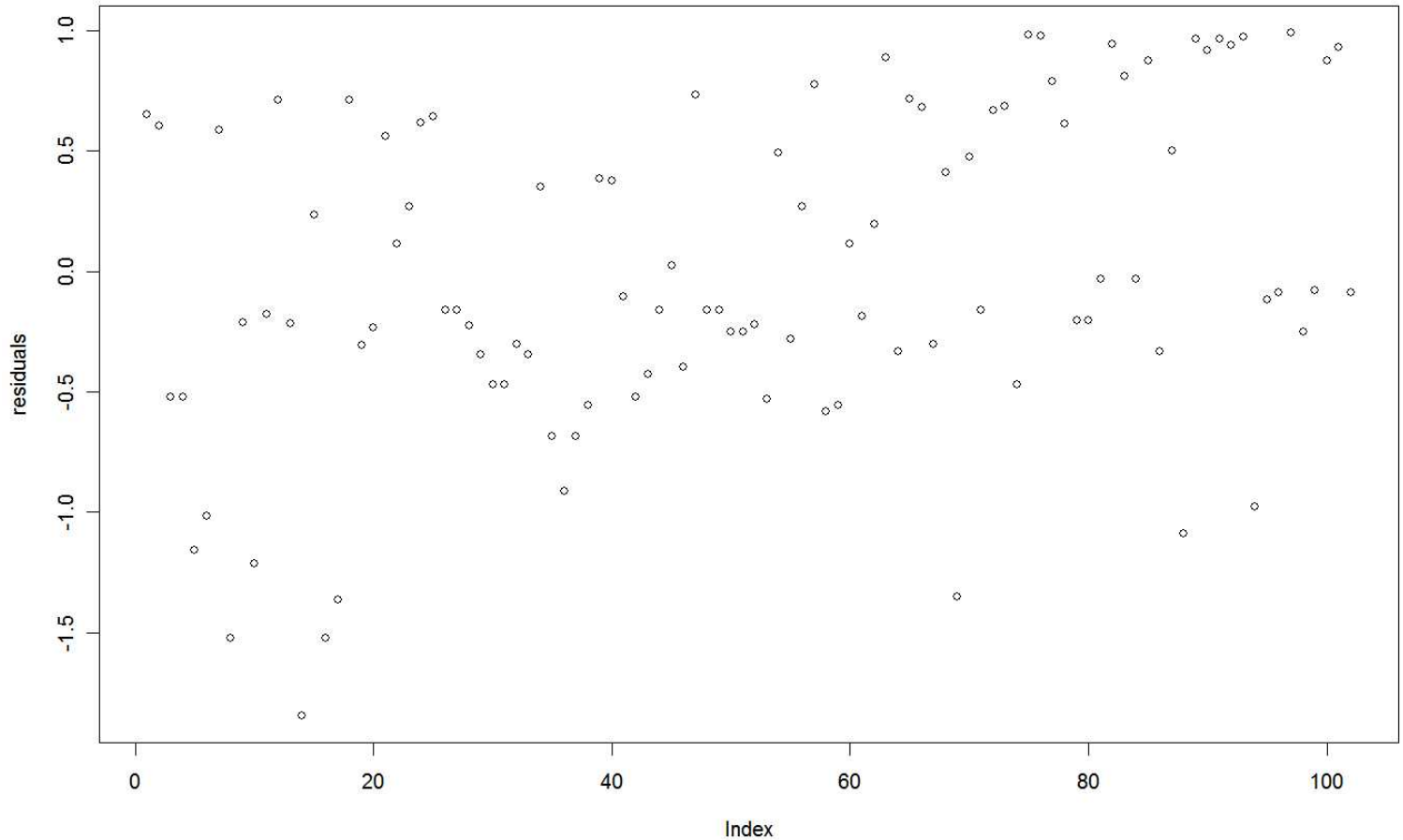
We note that this leads to a multivariate residual.

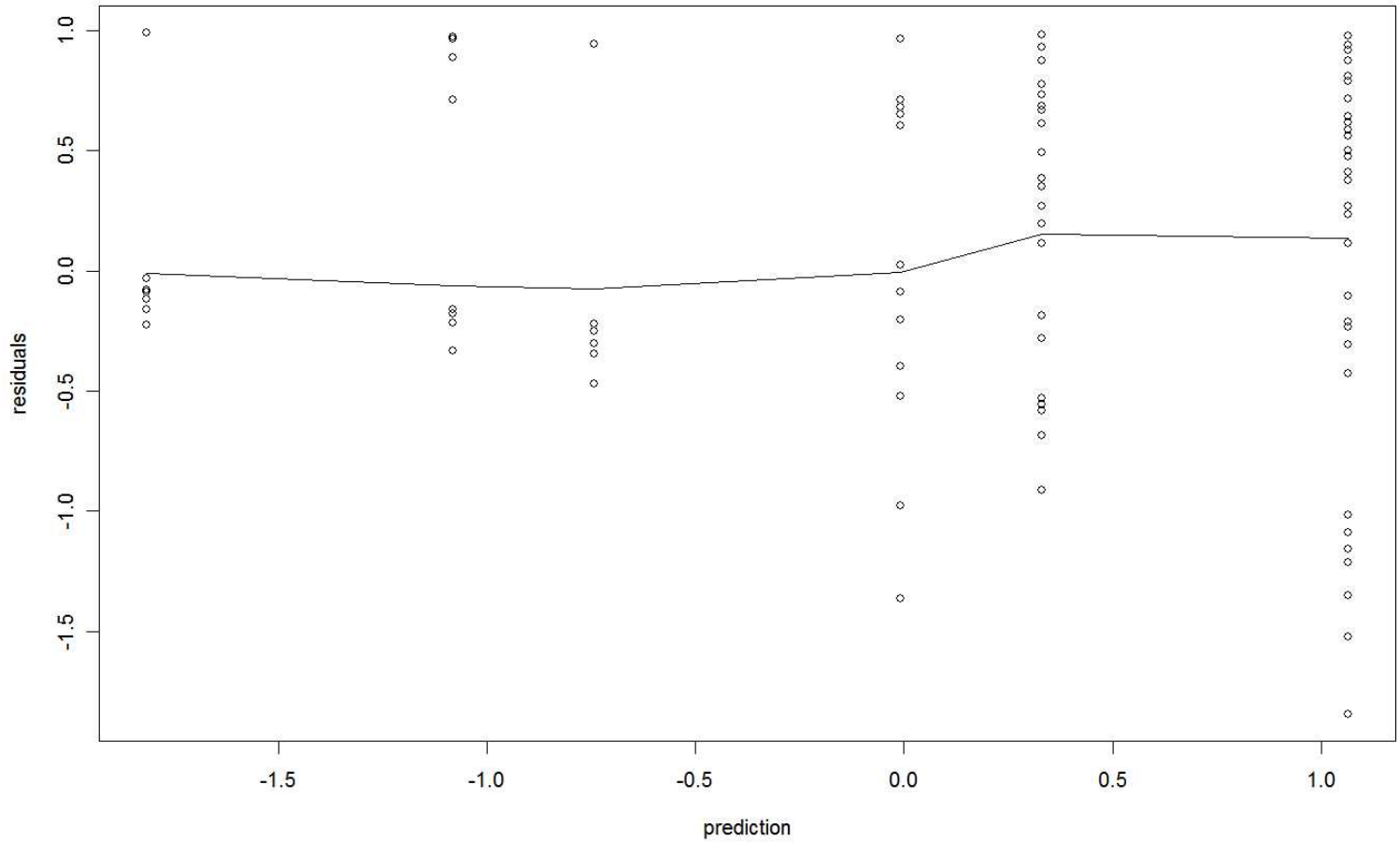
We use these residuals to asses whether the covariates satisfy the proportional hazards assumption.

Hereto we plot them versus ranks of the survival times.

Example: NSCLC

```
> fit<-coxph(Surv(survtime,survind)~expres+tnm,data=NSCLC)
> summary(fit)
>
> fitresi<-residuals(fit,type="martingale")
> fitpred<-predict(fit)
>
> plot(fitresi,ylab="residuals")
>
> plot(fitpred,fitresi,xlab="prediction",ylab="residuals")
> lines(lowess(fitpred,fitresi))
>
> plot(NSCLC$expres,fitresi,xlab="expression",ylab="residuals")
> lines(lowess(NSCLC$expres,fitresi))
>
> plot(NSCLC$tnm,fitresi,xlab="TNM",ylab="residuals")
> lines(lowess(NSCLC$tnm,fitresi))
```


residuals
1.0
0.5
0.0
-0.5
-1.0
-1.5
0
20
40
60
80
100
Index


residuals
1.0
0.5
0.0
-0.5
-1.0
-1.5
-1.5
-1.0
-0.5
0.0
0.5
1.0
prediction

Until now, we assumed that a covariate X satisfied the proportional hazards assumption. However in practice we need to check for this.

Under this assumption, we have that

$$\Lambda(t|X) = \int_0^t \lambda(s|X)ds = e^{\beta X} \int_0^t \lambda_0(s)ds = e^{\beta X}\Lambda_0(t)$$

and

$$S(t|X) = e^{-\Lambda(t|X)} = e^{-\Lambda_0(t)e^{\beta X}} = S_0(t)^{e^{\beta X}}.$$

Hence,

$$\log(-\log(S(t|X))) = \log(-\log(S_0(t))) + \beta X.$$

We use this relationship for a graphical check of proportional hazards.

Consider a discrete covariate X with levels $x_1, \ldots, x_K$.

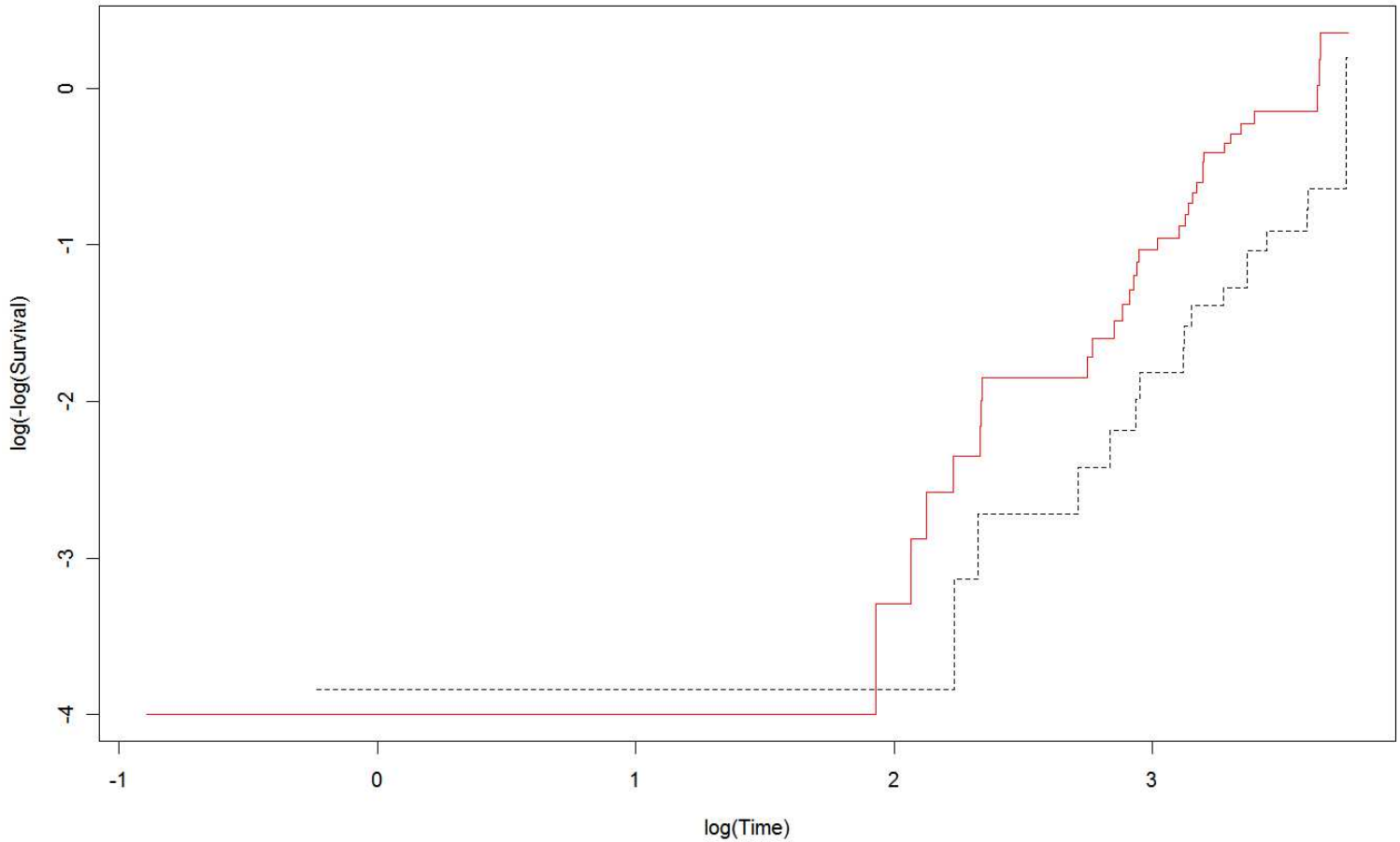
- Calculate KM curves for the various levels of X.
- Plot $\log(-\log(S(t|X = x))$ versus $\log(t)$.
- If they are parallel, the proportional hazard assumption is satisfied.

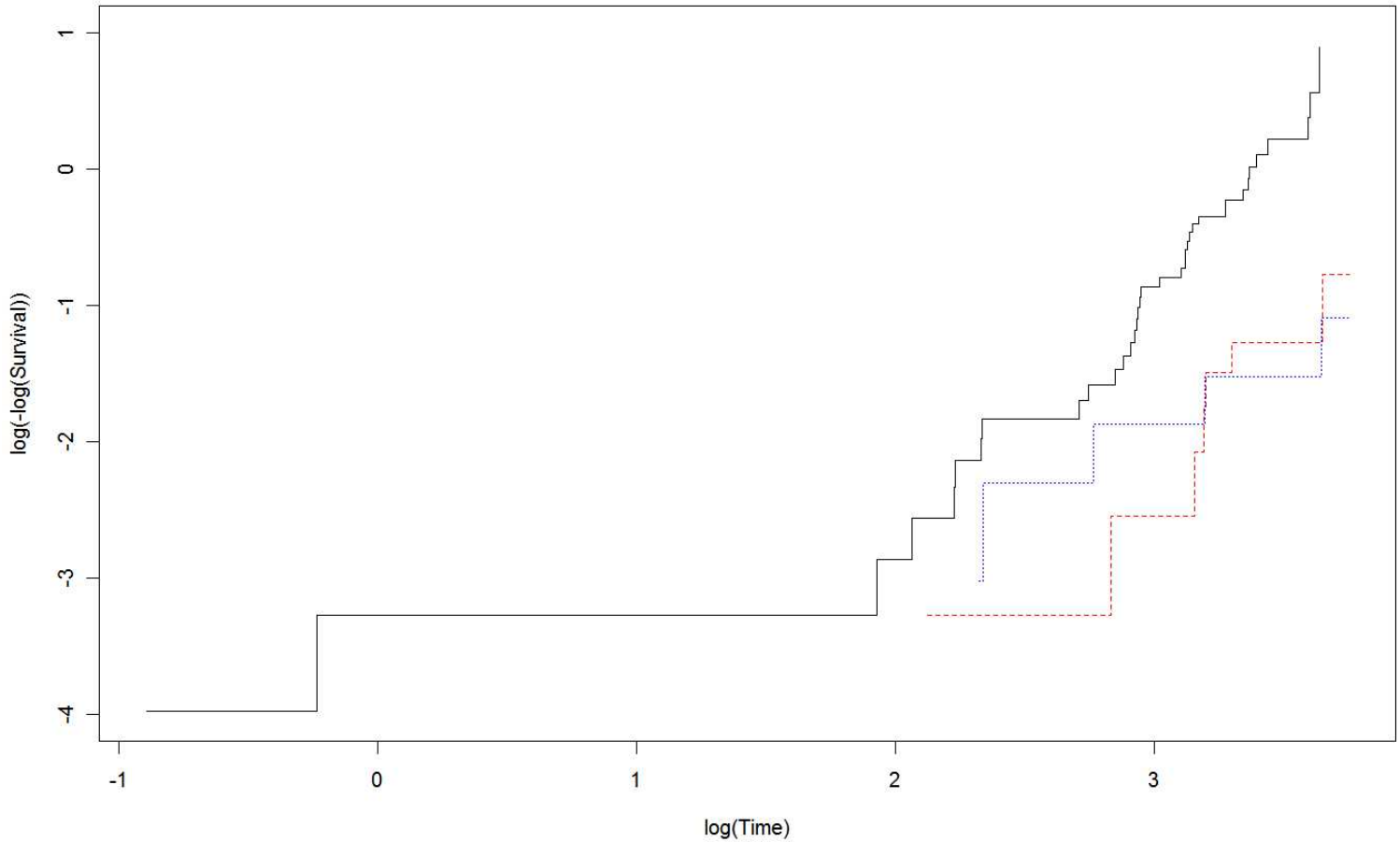
Note:

- For a continuous variable, we discretizes into categories.
- For more than one variable, we calculate a KM curve for each combination of values.

```
fit<-survfit(Surv(survtime,survind)~expres,data=NSCLC)
summary(fit)
plot(log(fit[2]$time),log(-log(fit[2]$surv)),xlab="log(Time)",ylab="log(-log(Survival)",
type="s",col="red")
lines(log(fit[1]$time),log(-log(fit[1]$surv)),lty=2,type="s")

fit<-survfit(Surv(survtime,survind)~tnm,data=NSCLC)
summary(fit)
plot(log(fit[3]$time),log(-log(fit[3]$surv)),xlab="log(Time)",ylab="log(-log(Survival))",
type="s",ylim=c(-4,1))
lines(log(fit[2]$time),log(-log(fit[2]$surv)),lty=2,type="s",col="red")
lines(log(fit[1]$time),log(-log(fit[1]$surv)),lty=3,type="s",col="blue")
```


log(-log(Survival))
0
-1
-2
-3
-4
-1
0
1
2
3
log(Time)


log(-log(Survival))
1
0
-1
-2
-3
-4
-1
0
1
2
3
log(Time)

Until now we assumed only one covariate but the idea also holds for multiple covariates.

Consider a discrete covariate X with levels $x_1, \ldots, x_K$.

- Calculate the fitted curves containing other covariates for the various levels of X.
- Plot $\log(-\log(S(t|X = x))$ versus $\log(t)$.
- If they are parallel, the proportional hazard assumption is satisfied.

A major use of time-dependent covariate methodology is to test the proportionality assumption for a fixed-time covariate X_1.

- Define an **artificial** time-dependent covariate $X_2(t)$,

$$X_2(t) = X_1 \times g(t)$$

  where $g(t)$ is a known function of time t (ex: $g(t) = \log(t)$).
- Test $H_0 : \beta_2 = 0$ in the Cox's model,

$$\lambda(t|X_1, X_2(t)) = \lambda_0(t)e^{\beta_1 X_1 + \beta_2 X_2(t)}.$$

Now,

$$HR(t) = e^{\beta_1(X_1 - X_1^*) + \beta_2 g(t)(X_1 - X_1^*)}$$

which is independent of t when $\beta_2 = 0$.

Example: NSCLC

```
> fit<-coxph(Surv(survtime,survind)~expres+tnm+tt(expres),data=NSCLC,tt=function(x,t,...)x*log(t))
> summary(fit)
Call: coxph(formula = Surv(survtime, survind) ~ expres + tnm + tt(expres),data = NSCLC,
 tt = function(x, t, ...) x * log(t))

n= 102, number of events= 49

              coef exp(coef) se(coef)     z Pr(>|z|)
expres     0.49387   1.63864  1.08088 0.457    0.648
tnm        1.07384   2.92660  0.25485 4.214 2.51e-05 ***
tt(expres) 0.08354   1.08712  0.36139 0.231    0.817
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

           exp(coef) exp(-coef) lower .95 upper .95
expres         1.639     0.6103    0.1970    13.631
tnm            2.927     0.3417    1.7760     4.823
tt(expres)     1.087     0.9199    0.5354     2.207

Concordance= 0.7  (se = 0.287 )
Rsquare= 0.259   (max possible= 0.98 )
Likelihood ratio test= 30.63  on 3 df,   p=1.015e-06
Wald test            = 23.79  on 3 df,   p=2.761e-05
Score (logrank) test = 27.03  on 3 df,   p=5.792e-06
```

```
> fit<-coxph(Surv(survtime,survind)~expres+tnm+tt(tnm),data=NSCLC,tt=function(x,t,...)x*log(t))
> summary(fit)
Call: coxph(formula = Surv(survtime, survind) ~ expres + tnm + tt(tnm),data = NSCLC,
 tt = function(x, t, ...) x * log(t))

n= 102, number of events= 49

          coef exp(coef) se(coef)     z Pr(>|z|)
expres  0.7399    2.0957   0.3035 2.438   0.0148 *
tnm     0.5151    1.6737   0.7837 0.657   0.5110
tt(tnm) 0.1917    1.2113   0.2612 0.734   0.4632
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

        exp(coef) exp(-coef) lower .95 upper .95
expres      2.096     0.4772    1.1560     3.799
tnm         1.674     0.5975    0.3603     7.776
tt(tnm)     1.211     0.8256    0.7259     2.021

Concordance= 0.695  (se = 0.287 )
Rsquare= 0.262   (max possible= 0.98 )
Likelihood ratio test= 31.05  on 3 df,   p=8.303e-07
Wald test            = 24.22  on 3 df,   p=2.251e-05
Score (logrank) test = 27.67  on 3 df,   p=4.254e-06
```

Example: NSCLC

```
proc phreg data=nsclc;
model Survtime*Survind(0)=expres TNM exptime;
exptime=expres*log(Survtime);
run;
```

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
expres	1	0.49389	1.08088	0.2088	0.6477	1.639
tnm	1	1.07358	0.25482	17.7503	<.0001	2.926
exptime	1	0.08352	0.36139	0.0534	0.8172	1.087

```
proc phreg data=nsclc;
model Survtime*Survind(0)=expres TNM TNMtime;
TNMtime=TNM*log(Survtime);
run;
```

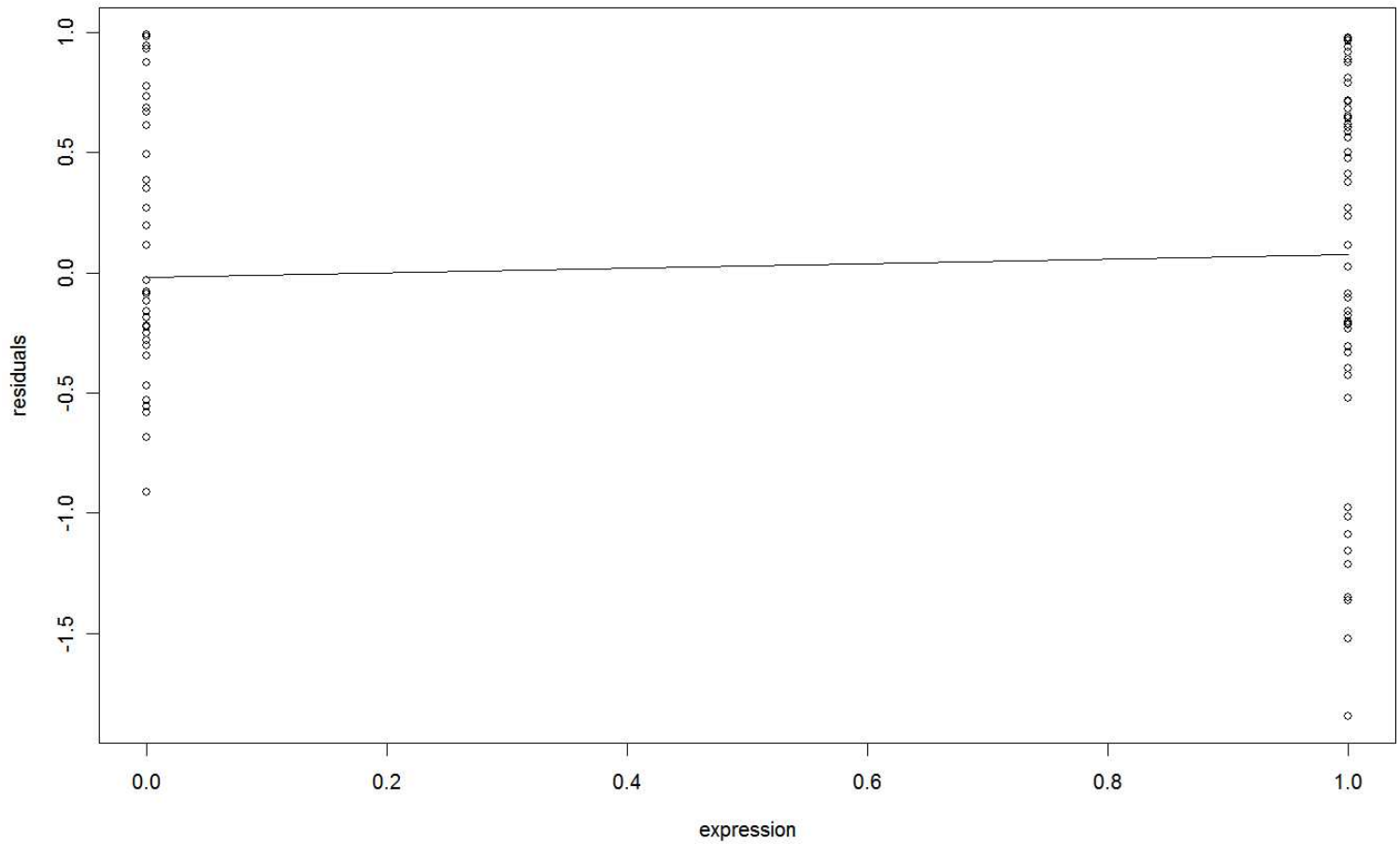
Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
expres	1	0.73986	0.30352	5.9417	0.0148	2.096
tnm	1	0.51472	0.78357	0.4315	0.5113	1.673
TNMtime	1	0.19170	0.26120	0.5387	0.4630	1.211

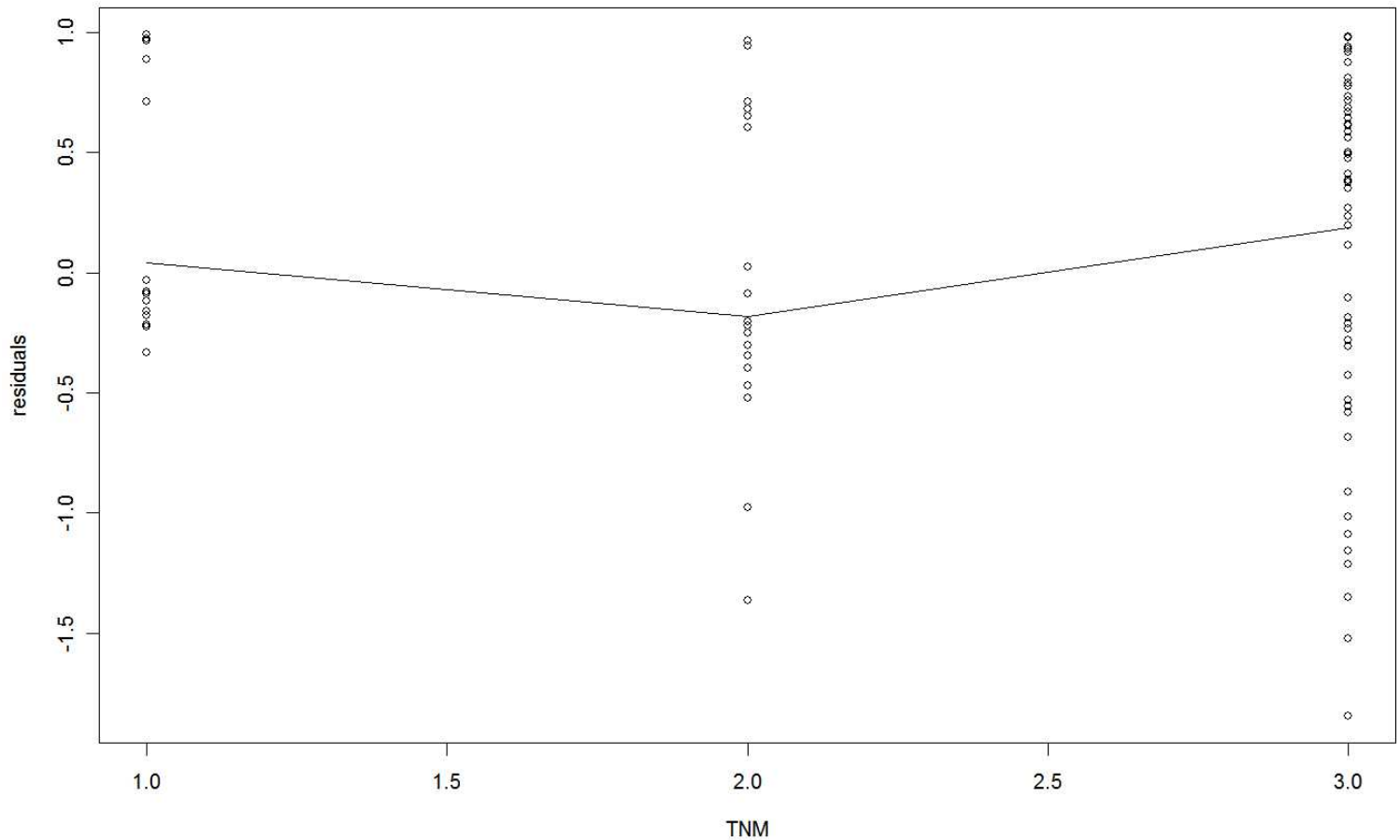
Sometimes, we are interested in checking which the best functional form of a covariate.

$$\lambda(t|\mathbf{X}) = \lambda_0(t)\exp(f(x_i)\beta_1 + x_m^t\beta)$$

To check which functional form is best, we can estimate several functional forms and compare them by the partial likelihood ratio test.

By plotting the martingale residuals versus a covariate, we can verify whether the functional form is correct.


residuals
1.0
0.5
0.0
-0.5
-1.0
-1.5
0.0
0.2
0.4
0.6
0.8
1.0
expression


residuals
1.0
0.5
0.0
-0.5
-1.0
-1.5
1.0
1.5
2.0
2.5
3.0
TNM