TMA4275 LIFETIME ANALYSIS

Slides 12: Weibull regression; Cox regression

Bo Lindqvist
Department of Mathematical Sciences
Norwegian University of Science and Technology
Trondheim

http://www.math.ntnu.no/∼bo/
bo@math.ntnu.no

NTNU, Spring 2014

Special case of log-location-scale-survival-regression models.

Recall: If $T \sim \text{Weibull}(\theta, \alpha)$ then by definition

$$\begin{aligned} R(t) &= e^{-(\frac{t}{\theta})^{\alpha}} \\ z(t) &= \frac{\alpha t^{\alpha-1}}{\theta^{\alpha}} = \alpha \theta^{-\alpha} t^{\alpha-1} \\ \ln T &= \ln \theta + \frac{1}{\alpha} W, \text{ where } W \sim \text{Gumbel}(0,1) \end{aligned}$$

Weibull regression model for a lifetime T and corresponding covariate vector $\mathbf{x}$:

$$\ln T = \underbrace{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k}_{\ln \theta} + \frac{1}{\alpha} W = \underbrace{\beta_0 + \boldsymbol{\beta}' \mathbf{x}}_{\ln \theta} + \frac{1}{\alpha} W$$

Thus $\theta = e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k} \equiv e^{\beta_0 + \boldsymbol{\beta}' \mathbf{x}}$

Thus for Weibull regression for $(T, \mathbf{x})$,

$$T \sim \text{Weibull}(e^{\beta_0+\beta_1 x_1+\cdots+\beta_k x_k}, \alpha),$$

and hence the hazard rate function is

$$\begin{aligned} z(t;\mathbf{x}) &= \alpha(e^{\beta_0+\beta_1 x_1+\cdots+\beta_k x_k})^{-\alpha} t^{\alpha-1} \\ &= \underbrace{\alpha e^{-\alpha\beta_0} t^{\alpha-1}}_{z_0(t)} \cdot e^{-\alpha\beta_1 x_1+-\alpha\beta_2 x_2\cdots+-\alpha\beta_k x_k} \\ &= z_0(t)\cdot e^{\tilde\beta_1 x_1+\tilde\beta_2 x_2+\cdots+\tilde\beta_k x_k}; \quad \text{where} \quad \tilde\beta_j = -\alpha\beta_j \\ &= z_0(t)\cdot e^{\tilde{\boldsymbol\beta}'\mathbf{x}} \\ &= z_0(t)\cdot g(\mathbf{x}) \end{aligned}$$

Thus: Hazard rate is product of one factor, $z_0(t)$, which is a function of t (and not of $\mathbf{x}$), and one which is function of $\mathbf{x}$ (and not of t).
This property is called the **Proportional hazards property**. Why? (Next slide).

PROPORTIONAL HAZARDS PROPERTY (CONT.)

Recall that $z(t; \mathbf{x}) = z_0(t) \cdot g(\mathbf{x})$. Consider two individuals with covariate vectors $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$:

$$\frac{z(t; \mathbf{x}^{(1)})}{z(t; \mathbf{x}^{(2)})} = \frac{g(\mathbf{x}^{(1)})}{g(\mathbf{x}^{(2)})} \qquad (\star)$$

Thus

$$z(t; \mathbf{x}^{(1)}) = \frac{g(\mathbf{x}^{(1)})}{g(\mathbf{x}^{(2)})} z(t; \mathbf{x}^{(2)})$$

so the hazard rate functions are proportional as functions of t, with proportionality factor equal to $g(\mathbf{x}^{(1)})/g(\mathbf{x}^{(2)})$.

Thus: The Weibull regression model has the proportional hazards property. **BUT no other** log-location-scale-survival-regression model has the property.

$(\star)$ is called the *relative risk* for a "person" with covariate $\mathbf{x}^{(1)}$ relative to a "person" with $\mathbf{x}^{(2)}$.

Sir David Cox, in his famous paper from 1972 suggested to use the model

$$z(t;\mathbf{x}) = z_0(t)e^{\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k}$$

Here $z_0(t)$ can be *any* positive function of t (i.e. any nonparametric hazard rate function). Because the $\beta_1, \ldots \beta_k$ are ordinary *parameters*, the model is said to be *semi-parametric*.

Interest is mainly in $\beta_1, \cdots, \beta_k$.

How to interpret β_i? Suppose an item has covariate vector $\mathbf{x} = (x_1, \cdots, x_k)$, so $z(t;\mathbf{x}) = z_0(t)e^{\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k}$. Suppose then that x_i (e.g. temperature) is increased by 1 unit, so $\mathbf{x}_{\text{new}} = x_1, \cdots, x_i + 1, \cdots, x_k$. Then

$$z(t;\mathbf{x}_{\text{new}}) = z(t \mid \mathbf{x}) \cdot e^{\beta_i}$$

Thus: e^{β_i} is the factor with which the hazard is multiplied if we increase X_i by 1 unit.

EXAMPLE - SMOKING OR NOT SMOKING

Suppose that the first component, x_1, of $\mathbf{x}$ is either 0 or 1:

- $x_1 = 0$ if person is *not* smoking.
- $x_1 = 1$ if person is smoking.

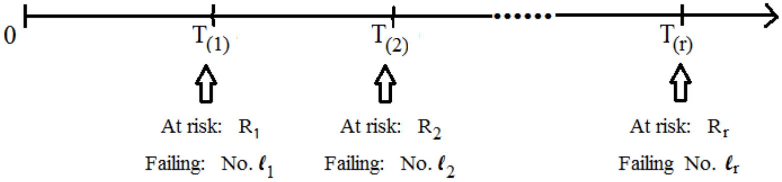
Then e^{β_1} is the effect on hazard rate caused by going from non-smoking to smoking, called the relative risk for a smoker.

In general: e^{β_i} is called the *relative risk* of covariate $\#i$.

Data: $(Y_i, \delta_i, \mathbf{X}_i),\ i = 1, \ldots, n$

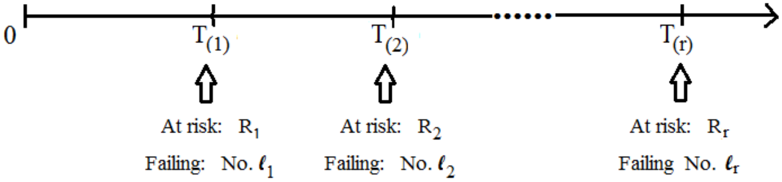
Model: $z(t; \mathbf{x}) = z_0(t) e^{\beta' \mathbf{x}}$

Let $T_{(1)} < T_{(2)} < \cdots < T_{(r)}$ be the observed *failure* times.



Need to know

- *who* are at risk at time $T_{(i)}$? Denote these $R_i \subseteq \{1, 2, \ldots, n\}$
- *who* fails at $T_{(i)}$? Say, this is individual $\ell_i \in R_i$.



Cox noted that since $z_0(t)$ is completely unknown, the lengths of times between failures are not relevant for estimation of β.

Cox' *partial likelihood* is essentially the likelihood of the observed $\ell_1, \cdots, \ell_k$:

$$L(\beta) = P(L_1 = \ell_1, L_2 = \ell_2, \cdots, L_k = \ell_k)$$

where L_i is the number of the indiviudaul that fails at time $T_{(i)}$.

Cox computed this as a product of the relevant probabilities at each failure time.

At $T_{(j)}$ there is a competition between all individuals in R_j, so we need to find in general, when t is one of $T_{(1)}, \dots, T_{(r)}$,

$$P(\ell_j \text{ fails at } t \mid \text{a unit in } R_j \text{ fails at } t)$$

$$= \frac{P(\ell_j \text{ fails at } t)}{P(\text{a unit in } R_j \text{ fails at } t)} \approx \frac{P(\ell_j \text{ fails in } (t, t+h))}{P(\text{a unit in } R_j \text{ fails in } (t, t+h))}$$

$$\approx \frac{z_0(t) e^{\boldsymbol{\beta}' \mathbf{x}_{\ell_j}} \cdot h}{\sum_{i \in R_j} z_0(t) e^{\boldsymbol{\beta}' \mathbf{x}_i} \cdot h} = \frac{e^{\boldsymbol{\beta}' \mathbf{x}_{\ell_j}}}{\sum_{i \in R_j} e^{\boldsymbol{\beta}' \mathbf{x}_i}}$$

so

$$L(\boldsymbol{\beta}) = \prod_{j=1}^{r} \frac{e^{\boldsymbol{\beta}' \mathbf{x}_{\ell_j}}}{\sum_{i \in R_j} e^{\boldsymbol{\beta}' \mathbf{x}_i}}$$

which is Cox' partial likelihood.

The log partial likelihood is $l(\boldsymbol{\beta}) = \ln L(\boldsymbol{\beta})$. The maximum partial likelihood estimate of $\boldsymbol{\beta}$ is found by solving

$$\frac{\partial l(\boldsymbol{\beta})}{\partial \beta_i} = 0; \; i = 1, \cdots, k$$

giving $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k)$, and in the same way as for parametric regression models,

$$I^{-1}(\hat{\boldsymbol{\beta}}) = \begin{bmatrix} \widehat{Var\hat{\beta}_1} & \cdot & \cdot \\ \widehat{cov(\hat{\beta}_1, \hat{\beta}_2)} & \widehat{Var\hat{\beta}_2} & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \widehat{Var\hat{\beta}_k} \end{bmatrix}$$

Assume d_j units fail at $T_{(j)}$. Peto-Breslow's partial likelihood:

$$L(\boldsymbol{\beta}) = \prod_{j=1}^{r} \frac{e^{\boldsymbol{\beta}' s_j}}{\left(\sum_{i \in R_j} e^{\boldsymbol{\beta}' \mathbf{x}_i}\right)^{d_j}}$$

where s_j is sum of $\mathbf{x}_\ell$ for the units that fail at $T_{(j)}$.

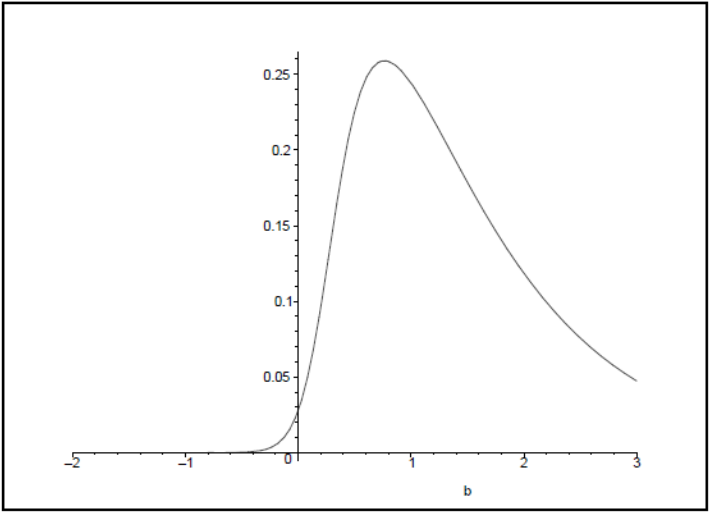
Essentially, we use Cox' partial likelihood by making an ordinary product for each failed unit, but we let all units that fail at the same time have the same risk set.

Model: $z(t;x) = z_0(t)e^{\beta x}$ (a single covariate, x).
Data: $n = 7$, $r = 3$

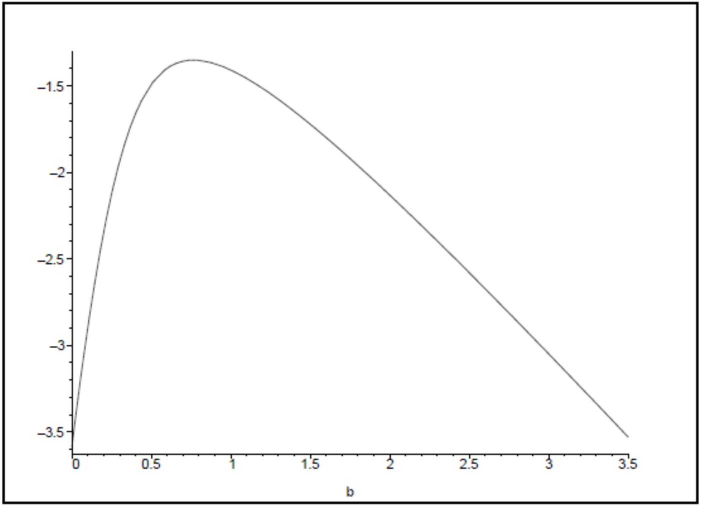
i	y_i	x_i	δ_i
1	5	12	0
2	10	10	1
3	40	3	0
4	80	5	0
5	120	3	1
6	400	4	1
7	600	1	0

j	$T_{(j)}$	R_j	ℓ_j
1	10	$\{2,3,4,5,6,7\}$	2
2	120	$\{5,6,7\}$	5
3	400	$\{6,7\}$	6

$$L(\beta) = \frac{e^{10\beta}}{e^{10\beta}+e^{3\beta}+e^{5\beta}+e^{3\beta}+e^{4\beta}+e^{\beta}} \cdot \frac{e^{3\beta}}{e^{3\beta}+e^{4\beta}+e^{\beta}} \cdot \frac{e^{4\beta}}{e^{4\beta}+e^{\beta}}$$



Maximum likelihood estimate: $\hat{\beta} = 0.765$.



Maximum likelihood estimate: $\hat{\beta} = 0.765$.

95% likelihood confidence interval: (0.1, 3.2).

Likelihood theory holds for the partial likelihood

$$W(\beta) = 2(l(\hat{\beta}) - l(\beta)) \approx \chi_1^2 \text{ if } \beta \text{ is true value.}$$

Thus we can construct the "1.92 Confidence Interval" (see previous slide), i.e. finding the set $\{\beta : l(\beta)) \geq l(\hat{\beta}) - 1.92\}$.

We can also test, e.g., $H_0 : \beta = 0$ versus $H_1 : \beta \neq 0$ by using that then

$$W = 2(l(\hat{\beta}) - l(0)) \sim \chi_1^2$$

under the null hypothesis, and reject H_0 if this becomes too big (larger than 3.84 for 5% significance level).

In example: $W = 2(-1.35 - (-3.45)) = 2 \cdot 2.10 = 4.2$, so we reject H_0 at 5% level. We could also conclude this from the confidence interval, since 0 is not in the confidence interval $(0.1, 3.2)$.

Distribution: Weibull

Relationship with accelerating variable(s): Linear

Regression Table

Predictor	Coef	Standard Error	Z	P	95,0% Normal CI Lower	95,0% Normal CI Upper
Intercept	7,58636	0,548229	13,84	0,000	6,51185	8,66087
x	-0,468235	0,0842830	-5,56	0,000	-0,633427	-0,303044
Shape	2,05563	0,872169			0,894943	4,72167

Log-Likelihood = -17,450

Worksheet 1 ***

↓	C1	C2	C3	C4	C5	C6	C7	C8
	Y	x	d					
1	5	12	0					
2	10	10	1					
3	40	3	0					
4	80	5	0					
5	120	3	1					
6	400	4	1					
7	600	1	0					

Estimated model, Weibull: $\ln T = 7.586 - 0.468x + (1/2.056)W$

Estimated model, Cox: $z(t;x) = z_0(t)e^{0.765x}$

Recall from earlier slide:

$$\beta_{cox} = -\alpha_{weib} \cdot \beta_{weib}$$

In the example we estimate the right hand side by $-2.056 \cdot (-0.468) = 0.96$ while the left hand side is estimated by 0.765.

This seems to be OK, given that there are very few failures, and given the following fact:

The Cox-estimate for β does not use the observed times, while the Weibull estimates use them (a lot).

Table 3.2. Lifetimes (in cycles) of sodium sulphur batteries

Batch 1	164	164	218	230	263	467	538	639	669
	917	1148	1678+	1678+	1678+	1678+			
Batch 2	76	82	210	315	385	412	491	504	522
	646+	678	775	884	1131	1446	1824	1827	2248
	2385	3077							

Note: Lifetimes with + are right censored observations, not failures.

There are altogether $n = 15 + 20 = 35$ observations.

Let $x = 0$ for Batch 1, $x = 1$ for Batch 2.

Now x is a discrete covariate (categorical). The Cox model is $z(t; x) = z_0(t)e^{\beta x}$, so:

- for Batch 1: $z(t; 0) = z_0(t)$
- for Batch 2 : $z(t; 1) = z_0(t)e^{\beta}$

Cox' partial likelihood is easy to write down here (but note tied failures at time 164, so Peto-Breslow should be used at that time). For the other times, the contribution at $T(j)$ is

$$\frac{e^{\boldsymbol{\beta}' \mathbf{x}_{\ell_j}}}{\sum_{i \in R_j} e^{\boldsymbol{\beta}' \mathbf{x}_i}} = \frac{1 \text{ if failure in Batch 1, } e^{\beta} \text{ if failure in Batch 2}}{\#\text{at risk in Batch 1} + e^{\beta} \cdot \#\text{at risk in Batch 2}}$$

and Cox' likelihood is the product of these!

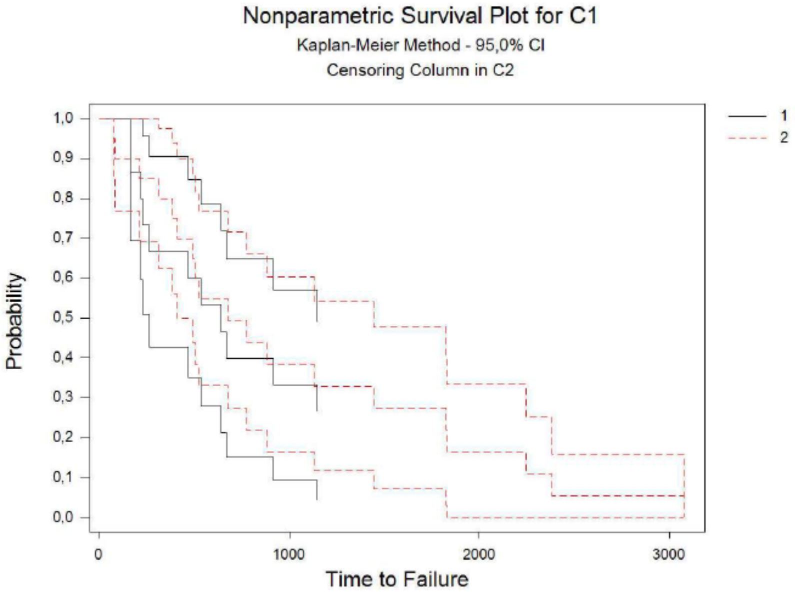
Maximum partial likelihood estimate: $\hat{\beta} = -0.0888$
(solved $\frac{\partial l(\beta)}{\partial \beta} = 0$, where l is Cox' log partial likelihood)

Further, computation of $\widehat{Var(\hat{\beta})} = (-l''(\hat{\beta}))^{-1}$, and taking the square root gives the standard error $\widehat{SD(\hat{\beta})} = 0.4034$.

So the standard 95% confidence interval for β is $-0.0888 \pm 1.96 \cdot 0.4034$
$= (-0.879, 0.702)$.

To test $H_0 : \beta = 0$ versus $H_1 : \beta \neq 0$
use $W(0) = 2(l(\hat{\beta}) - l(0)) \approx \chi^2_1$ under H_0
$= 2(-81.238 - (-81.262)) = 0.048$ so do not reject at any reasonable significance level!

Note that we could also use the logrank test to test these hypotheses, or look at KM-plots (next slide).


Nonparametric Survival Plot for C1
Kaplan-Meier Method - 95,0% CI
Censoring Column in C2
Probability
1,0
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0,0
0
1000
2000
3000
Time to Failure
1
2

- Hazard: $z(t; \mathbf{x}_i) = z_0(t) e^{\boldsymbol{\beta}' \mathbf{x}_i}$
- Cumulative hazard: $Z(t; \mathbf{x}_i) = Z_0(t) e^{\boldsymbol{\beta}' \mathbf{x}_i}$
  (do the integration!)
- Survival/reliability function:

  $$P(T_i > t) = R(t; \mathbf{x}_i) = e^{-Z(t; \mathbf{x}_i)} = e^{-Z_0(t) e^{\boldsymbol{\beta}' \mathbf{x}_i}}$$

Of practical interest: "Estimate the survival probability for a patient or machine".

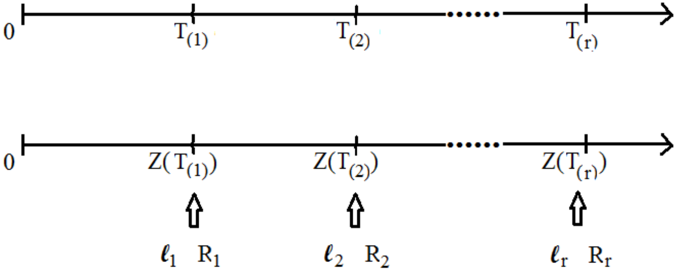
To estimate this: Substitute $\hat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}$, but still we need to estimate $Z_0(t)$.

Recall: $Z(T_i, \mathbf{x}_i) \sim \text{expon}(1)$, i.e. $Z_0(T_i)e^{\boldsymbol{\beta}'\mathbf{x}_i} \sim \text{expon}(1)$

Then recall that $T \sim \text{expon}(\lambda) \Longrightarrow aT \sim \text{expon}(\lambda/a)$. But then

$Z_0(T_i) \sim \text{expon}(\underbrace{e^{\boldsymbol{\beta}'\mathbf{x}_i}}_{\lambda_i \text{for simplicity}})$, since

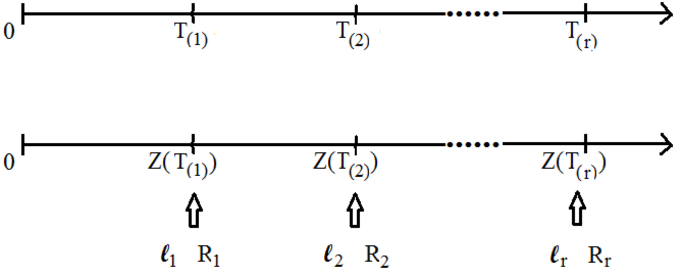
$$Z_0(T_i) = e^{-\boldsymbol{\beta}'\mathbf{x}_i} \cdot \underbrace{Z_0(T_i)e^{\boldsymbol{\beta}'\mathbf{x}_i}}_{\text{expon}(1)} \sim \text{expon}\left(\frac{1}{e^{-\boldsymbol{\beta}'\mathbf{x}_i}}\right) = \text{expon}(e^{\boldsymbol{\beta}'\mathbf{x}_i})$$



$Z_0(T_i) \sim \text{expon}(e^{\boldsymbol{\beta}' \mathbf{x}_i}) \equiv \text{expon}(\lambda_i)$, so

$Z_0(T_{(1)}) = \text{minimum of } Z_0(T_i) \text{ for } i \in R_1 \sim \text{expon}(\sum_{i \in R_1} \lambda_i) =$
$\text{expon}(\sum_{i \in R_1} e^{\boldsymbol{\beta}' \mathbf{x}_i})$, and

$Z_0(T_{(2)}) - Z_0(T_{(1)}) \sim \text{expon}(\sum_{i \in R_2} \lambda_i) = expon(\sum_{i \in R_2} e^{\boldsymbol{\beta}' \mathbf{x}_i})$



It follows that $E(Z_0(T_{(1)})) = \frac{1}{\sum_{i \in R_1} \lambda_i}$

$$E(Z_0(T_{(2)})) = \frac{1}{\sum_{i \in R_1} \lambda_i} + \frac{1}{\sum_{i \in R_2} \lambda_i}$$

and so on, so that in general

$$E(Z_0(T_{(m)})) = \sum_{j=1}^{m} \frac{1}{\sum_{i \in R_j} \lambda_i} = \sum_{j=1}^{m} \frac{1}{\sum_{i \in R_j} e^{\hat{\beta}' x_i}}$$

$$\hat{Z}_0(t) = \sum_{T_{(j)} \leq t} \frac{1}{\sum_{i \in R_j} e^{\hat{\beta}' \mathbf{x}_i}}$$

This is similar to Nelson-Aalen estimator.

Indeed, if there are no covariates, then $\beta = 0$ and we get $\sum_{T_j \leq t} \frac{1}{\# R_j}$, which is the Nelson-Aalen estimator.

We can use the Breslow estimator to estimate $\hat{R}_0(t) = e^{-\hat{Z}_0(t)}$.

$$\hat{R}_0(t) = \prod_{j:T_{(j)} \le t} \left(1 - \frac{e^{\hat{\boldsymbol{\beta}}' \mathbf{x}_{l_j}}}{\sum_{i \in R_j} e^{\hat{\boldsymbol{\beta}}' \mathbf{x}_i}}\right)^{e^{-\hat{\boldsymbol{\beta}}' \mathbf{x}_{l_j}}}$$

Note that for $\boldsymbol{\beta} = 0$ we get the ordinary KM estimator:

$$\hat{R}_0(t) = \prod_{j:T_{(j)} \le t} \left(1 - \frac{1}{\#R_i}\right)$$