Chapter 6. Modelling method for survival analysis

Type of Regression Models:

(1) Proportional Hazards Model (Cox Model)

$$h(t|z) = h_0(t)\Phi(z)$$

where $\Phi(z)$ is a positive function with $\Phi(0) = 1$ (ie. $h(t|z = 0) = h_0(t)$). In addition, if T is continuous then Cox model can also be specified by $S(t|z) = [S_0(t)]^{\Phi(z)}$. (Note: $S(t|z) = \exp\{-\int_0^t h(u|z)du\} = \exp\{-\int_0^t h_0(u)\Phi(z)du\} = [S_0(t)]^{\Phi(z)}$)

Usually, we define $\Phi(z) = \exp\{\beta' z\}$ where β is a p dimension unknown parameters that relates the covariate z to survival. It can be interpreted as relative risk since

$$\frac{h(t|z)}{h(t|z = 0)} = \exp\{\beta' z\}.$$

Example 1: $z = (1, z_1, \ldots, z_{p-1})'$ and $\beta = (\alpha, \beta_1, \ldots, \beta_{p-1})'$

$$\begin{aligned} h_0(t)\exp\{\beta' z\} &= h_0(t)\exp\{\alpha + \sum_{i=1}^{p-1}\beta_i z_i\} \\ &= h_0^*(t)\exp\{\beta^{*\prime} z^*\} \end{aligned}$$

where $z^* = (z_1, \ldots, z_{p-1})'$ and $\beta^* = (\beta_1, \ldots, \beta_{p-1})'$.

Example 2: Suppose $z = \begin{cases} 0, & \text{if treatment} \\ 1, & \text{if control} \end{cases}$, then $h(t|z) = \begin{cases} h_0(t)\exp\{\beta\}, & \text{if } z = 1 \\ h_0(t), & \text{if } z = 0 \end{cases}$.
Thus to test $S_1 = S_0 \Leftrightarrow$ to test $\beta = 0$.

(2) Constant Excess Risk Model

$$h(t|z) = h_0(t) + \Phi(z)$$

where $\Phi(z)$ is a positive function with $\Phi(0) = 1$. It reduces to additive hazards model when $\Phi(z) = \beta' z$.

(3) Accelerated Life Model

$$S(t|z) = S_0(\Phi(z)t).$$

Again, $\Phi(z)$ is a positive function with $\Phi(0) = 1$. Commonly, $\Phi(z) = \exp\{\beta' z\}$ is used. Accelerated life model can also be specified as

$$f(t|z) = \Phi(z)f_0(\Phi(z)t) \text{ and } h(t|z) = \Phi(z)h_0(\Phi(z)t).$$

Furthermore, suppose $T_i|z_i \tilde{} S(t|z)$ and $T_0|z = 0 \tilde{} S_0(t)$, accelerated life model can also represented as

$$\ln T_i = \ln T_0 - \ln \Phi(z_i).$$

(Note: $S(t|z) = P(T_i \geq t|z_i) = P(\ln T_i \geq \ln t|z_i) = P(\ln T_0 \geq \ln t + \ln \Phi(z_i)) = P(T_0 \geq t\Phi(z_i)) = S_0(\Phi(z)t)$.)

Example 1: Suppose $z = \begin{cases} 0, & \text{if treatment} \\ 1, & \text{if control} \end{cases}$, then $S(t|z) = \begin{cases} S_0(t\exp\{\beta\}), & \text{if } z = 1 \\ S_0(t), & \text{if } z = 0 \end{cases}$.
As we can see, $t\exp\{\beta\} = 1 \Leftrightarrow t = \exp\{-\beta\}$. That is control lives 1 unit, treatment lives $\exp\{-\beta\}$ units.

Parametric Approach:

Assume $h_0(t)$ is a function from a parametric family. For an example, $h_0(t) = k$ if exponential distribution is assumed, $h_0(t) = kt^\alpha$ if Weibull distribution is used. Then, consider the method of maximum likelihood to get the estimate of the unknown parameters.

In the case of right censoring, suppose the n observations are (Y_i, δ_i, Z_i) where $Y_i = \min(T_i, C_i)$ and $\delta_i = I(T_i = Y_i)$. Then the likelihood function is

$$\begin{aligned} L(\cdot|Y, \delta, Z) &= \prod_{i=1}^{n} [f(Y_i|Z_i)]^{\delta_i} [S(Y_i|Z_i)]^{1-\delta_i} \\ &= \prod_{i=1}^{n} [h(Y_i|Z_i)]^{\delta_i} S(Y_i|Z_i). \end{aligned}$$

For an example, if we assume $h(t|z) = h_0(t)\exp\{\beta z\}$ and $h_0(t) = \lambda_0$, the corresponding log likelihood function becomes

$$\begin{aligned} l(\beta, \lambda|Y, \delta, Z) &= \sum_{i=1}^{n} [\delta_i(\ln \lambda + \beta Z_i) - \lambda Y_i \exp\{\beta Z_i\}] \\ &= r \ln \lambda + \beta \sum_{i=1}^{n} \delta_i Z_i - \lambda \sum_{i=1}^{n} Y_i \exp\{\beta Z_i\}. \end{aligned}$$

where $r = \sum_{i=1}^{n} \delta_i = \#$ of death. Suppose $z = \begin{cases} 0, & \text{if treatment} \\ 1, & \text{if control} \end{cases}$, then
(a) $\sum_{i=1}^{n} \delta_i Z_i = \#$ of death in treatment group $= r_1$;
(b) $\sum_{i=1}^{n} Y_i \exp\{\beta Z_i\} = \exp\{\beta\} W_1 + W_0$ where $W_1 =$ sum of follow-up time in treatment group and $W_0 =$ sum of follow-up time in control group.

This implies

$$l(\beta, \lambda|Y, \delta, Z) = r \ln \lambda + \beta r_1 - \lambda[\exp\{\beta\} W_1 + W_0].$$

Set $\begin{cases} \frac{\partial l}{\partial \beta} = r_1 - \lambda W_1 \exp\{\beta\} = 0 \\ \frac{\partial l}{\partial \lambda} = \frac{r}{\lambda} - [\exp\{\beta\} W_1 + W_0] = 0 \end{cases}$, we have $\begin{cases} \hat{\beta} = \ln[\frac{r_1/w_1}{r_0/w_0}] \\ \hat{\lambda} = r_0/w_0 \end{cases}$ where $r_0 = r - r_1$. Let λ_0 be the hazard for control group and λ_1 be the hazard for treatment group. According to the model assumption, we have $\begin{cases} \hat{\lambda}_0 = r_0/w_0 \\ \hat{\lambda}_1 = r_1/w_1 \end{cases}$ which give the same estimates as the MLE of exponential distribution (it is trivial, why?). On the other hand, we have $\hat{\lambda}_1/\hat{\lambda}_0 = \exp\{\hat{\beta}\}$ or identically $\hat{\beta} = \ln \hat{\lambda}_1 - \ln \hat{\lambda}_0$.

Preliminary of Partial likelihood (Urn model):

Suppose there is an urn which contains $n_0(t)$ white balls (control group) and $n_1(t)$ black balls (treatment group) at time t, respectively. Assume $h_0(t)$ be the "risk" for a white ball being picked up and $h_1(t)$ be the "risk" for a white ball being picked up. Let $h_1(t) = h_0(t)\exp\{\beta\}$. Suppose every ball in the urn has its own number. Let $Z_i = I($ the number i ball is black), then $n_0(t) = \sum_{i \text{ in urn}}(1 - Z_i)$ and $n_1(t) = \sum_{i \text{ in urn}} Z_i$. The "risk" can be rewritten as $h(t|Z) = h_0(t)\exp\{\beta Z\}$. Then the probability that the number 1 ball being picked up at time t is

$$\frac{h_0(t)\exp\{\beta Z_1\}}{\sum_{i \text{ in urn}} h_0(t)\exp\{\beta Z_i\}} = \frac{\exp\{\beta Z_1\}}{\sum_{i \text{ in urn}} \exp\{\beta Z_i\}}.$$

Let $R(t) = \{i|$ the number i ball in urn at time $t\}$. Suppose we pick up k balls at k different times $t_{(1)} < \cdots < t_{(k)}$. Denote the corresponding covariate Z of these k balls as $Z_{(1)}, \cdots, Z_{(k)}$. Here $Z_{(1)}$ mean the covariate of the first ball that we pick up and so on. Then, the likelihood function of the observation becomes

$$L(\beta) = \prod_{i=1}^{k} \frac{\exp\{\beta Z_{(i)}\}}{\sum_{j \in R(t_{(i)})} \exp\{\beta Z_j\}}.$$

Partial likelihood function

$$L(\beta) = \prod_{i=1}^{n} [\frac{\exp\{\beta Z_i\}}{\sum_{j \in R(y_i)} \exp\{\beta Z_j\}}]^{\delta_i} \text{ where } R(t) = \{j | y_j \geq t\}$$

Log partial likelihood function

$$l(\beta) = \sum_{i=1}^{n} \delta_i [\beta Z_i - \log \sum_{j \in R(y_i)} \exp\{\beta Z_j\}]$$

Partial score function

$$U(\beta) = \frac{\partial l(\beta)}{\partial \beta} = \sum_{i=1}^{n} \delta_i [Z_i - \bar{Z}_i]$$

where $\bar{Z}_i = \sum_{j \in R(y_i)} w_j Z_j$ and $w_j = \exp\{\beta Z_j\} / \sum_{j \in R(y_i)} \exp\{\beta Z_j\}$.
Sample information matrix

$$I(\beta) = \frac{-\partial U(\beta)}{\partial \beta} = \sum_{i=1}^{n} \delta_i [\frac{\sum_{j \in R(y_i)} Z_j^2 \exp\{\beta Z_j\}}{\sum_{j \in R(y_i)} \exp\{\beta Z_j\}} - \bar{Z}_i^2]$$

The estimator $\hat{\beta}$ of β based on partial likelihood function has the same asymptotic properties as regular MLE. Thus, in testing $H_0 : \beta = \beta_0$ vs $H_a : \beta \neq \beta_0$, we usually consider the following three tests.

(1) Wald's test

$$(\hat{\beta} - \beta_0)^T I(\beta_0)(\hat{\beta} - \beta_0) \tilde{} \chi_p^2$$

or

$$(\hat{\beta} - \beta_0)^T I(\hat{\beta})(\hat{\beta} - \beta_0) \tilde{} \chi_p^2$$

(2) Score test

$$U^T(\beta_0) I^{-1}(\beta_0) U(\beta_0) \tilde{} \chi_p^2$$

or

$$U^T(\beta_0) I^{-1}(\hat{\beta}) U(\beta_0) \tilde{} \chi_p^2$$

(3) Likelihood ratio test

$$2(l(\hat{\beta}) - l(\beta_0)) \tilde{} \chi_p^2$$

The following is a history review of partial likelihood estimation.

(1) 1972, Cox's original "conditional likelihood".

(2) Kalbfleisch and Prentice (1973) Marginal likelihood base on rank.

(3) Breslow (1974) Profile likelihood.

(4) Cox's "Partial likelihood"

Why is called "partial likelihood"?

Let T be the failure time and C be the censoring time. However, we can only observe (Y, δ) where $Y = \min(T, C)$ and $\delta = I(T \leq C)$. Suppose (Y_i, δ_i) $(i = 1, \cdots, n)$ be n generations of (Y, δ) and $t_1 < t_2 < \cdots < t_k$ be the k distinct failure time of the n observations (i.e assume no tie among failure time). Denote $t_0 = 0$ and $t_{k+1} = \infty$. Let b_j specify the censoring information in $[t_{j-1}, t_i)$ plus the information that one individual fails in the interval $[t_j, t_j + dt_j)$. Let a_j specify that item j fails in $[t_j, t_j + dt_j)$. It is trivial that $b_1, a_1, \cdots, b_k, a_k, b_{k+1}$ can be known when (Y_i, δ_i) $(i = 1, \cdots, n)$ are observed. However, (Y_i, δ_i) $(i = 1, \cdots, n)$ can not be known exactly if only $b_1, a_1, \cdots, b_k, a_k, b_{k+1}$ are given. Thus, the information of the sample will be reduced if we are using $b_1, a_1, \cdots, b_k, a_k, b_{k+1}$ rather than using (Y_i, δ_i) $(i = 1, \cdots, n)$. Fortunately, under the random censoring assumption, both of the samples contain the same information about the failure time. This can be imagined through the Kaplan-Meier estimator.

Let $a^{(j)} = \{a_1, \cdots, a_j\}$, $b^{(j)} = \{b_1, \cdots, b_j\}$ and $a^{(0)} = b^{(0)} = \{\}$ (an empty set). Then, the full likelihood function based on $b_1, a_1, \cdots, b_k, a_k, b_{k+1}$ is

$$\begin{aligned} L(\beta) &= f(b_1, a_1, \cdots, b_k, a_k, b_{k+1}|\beta) \\ &= \prod_{i=1}^{k} f(a_i|b^{(i)}, a^{(i-1)}, \beta) \prod_{j=1}^{k+1} f(b_j|b^{(j-1)}, a^{(j-1)}, \beta). \end{aligned}$$

Moreover, under Cox's model,

$$f(a_i|b^{(i)}, a^{(i-1)}, \beta) = \frac{\exp\{\beta Z_i\}}{\sum_{j \in R(y_i)} \exp\{\beta Z_j\}}.$$

If there are ties, SPSS still use

$$L(\beta) = \prod_{i=1}^{n} [\frac{\exp\{\beta Z_i\}}{\sum_{j \in R(y_i)} \exp\{\beta Z_j\}}]^{\delta_i}.$$

On the other hand, SAS consider "combining 2 balls into one ball". For example, if the first two individuals die at time y (sure, the corresponding $\delta = 1$) and

having covariates Z_1 and Z_2 respectively, then the probability to have such an event happened is

$$\frac{\exp\{\beta(Z_1+Z_2)\}}{\sum_{j,k\in R(y),j\neq k}\exp\{\beta(Z_j+Z_k)\}}.$$

To test whether a (full) model can be omitted to a (reduce) nested sub-model by using likelihood ratio test:

Suppose that the $p+q$ covariates $Z_1,\cdots,Z_p,Z_{p+1},\cdots,Z_{p+q}$ are fitted in the model (1)

$$h(t|z)=h_0(t)\exp\{\beta_1 z_1+\cdots+\beta_{p+q}z_{p+q}\}.$$

Consider a nested model of model (1) as model (2)

$$h(t|z)=h_0(t)\exp\{\beta_1 z_1+\cdots+\beta_p z_p\}.$$

Since model (1) has larger number of terms than model (2), model (1) must be a better fit to the data. This implies

$$l(\hat{\beta}_1,\cdots,\hat{\beta}_{p+q})\geq l(\hat{\beta}_1,\cdots,\hat{\beta}_p).$$

The statistical problem is then to determine whether the additional q terms in model (1) significantly improve the explanatory power of the model. If not, they might be omitted and model (2) would be adequate. The test that we usually used to this problem is LRT. When the sample size n is large enough, under the null hypothesis H_0 that the additional q terms can be ignored, we have

$$2[l(\hat{\beta}_1,\cdots,\hat{\beta}_{p+q})-l(\hat{\beta}_1,\cdots,\hat{\beta}_p)]\tilde{}\chi^2_q.$$

Here, the alternative hypothesis is that at least one term of the q terms is significantly improve the explanatory power of the model.

In addition, the null hypothesis H_0 is equivalent to $H_0^*:\beta_{p+1}=\cdots=\beta_{p+q}=0$. Thus, we can also consider the Wald's test. In this problem, Score test is difficult to do since the variance structure is too complicate even the asymptotic results. This is because $\beta_1,\cdots,\beta_p$ are still unknown under null hypothesis, if we plug in $\hat{\beta}_1,\cdots,\hat{\beta}_p$ into the score function, the asymptotic variance of the score function is not further the fisher information matrix.

For more general nested model testing problem, the asymptotic properties of LRT and the Wald's test are still hold. For an example, suppose the full model is

$$h(t|z)=h_0(t)\exp\{\beta_1 z_1+\beta_2 z_2+\beta_3 z_3+\beta_4 z_4\}.$$

If we are interested in knowing whether $H_0 : \beta_2 = \beta_3 = \beta_4 = \beta$, the reduce model becomes

$$h(t|z) = h_0(t) \exp\{\beta_1 z_1 + \beta(z_2 + z_3 + z_4)\}.$$

Then, under H_0, by using LRT

$$2[l(\text{Full}) - l(\text{Reduce})] \tilde{} \chi^2_2.$$

On the other hand, by using Wald's test, we can consider

$$(\hat{\beta}_2 - \hat{\beta}_3, \hat{\beta}_2 - \hat{\beta}_4)^T \hat{\Sigma}^{-1} (\hat{\beta}_2 - \hat{\beta}_3, \hat{\beta}_2 - \hat{\beta}_4) \tilde{} \chi^2_2,$$

where Σ is the covariance matrix of $(\hat{\beta}_2 - \hat{\beta}_3, \hat{\beta}_2 - \hat{\beta}_4)$. We can also consider some other equivalent test statistics, for example $(\hat{\beta}_3 - \hat{\beta}_2, \hat{\beta}_3 - \hat{\beta}_4)$ and so on. All of them will provide an identical p-value for the same data.

Model selection while the models are not nested:

If the models are not nested, the asymptotic properties of LRT is failed. A way to solve this problem is considering AIC (not recommend) where $AIC = -2l + \alpha p$ with $\alpha \in [2, 6]$ someone suggests using $\alpha = 3$. Here, p is the number of unknown parameters in the model. Then, model (1) is better than model (2) if the corresponding AIC of model (1) is smaller than that of model (2).

Note 1: In practical, the best way is to find the most reasonable and interpretable model.

Note 2: AIC does not consider the sample size. It might has good properties for smaller sample size n, however, for large n, it is poor. On the other hand, AIC might suggest a non-reasonable model.

Covariate selection procedure:

(i) Forward selection, (ii) Backward selection, (iii) Stepwise selection.

Interpretation:

Suppose

$$h(t|z) = h_0(t) \exp\{\beta z\}.$$

If z is dichotomize which takes value 0 or 1, then

$$\beta = \ln(RR)$$

where RR is the relative risk (hazard rate ratio). If z is continuous, then

$$\exp\{\beta\} = \frac{h(t|z+1)}{h(t|z)}.$$

That is, z increases one unit, the hazard rate ratio increase $\exp\{\beta\}$.

For an example:

Let $z_1 = \left\{ \begin{array}{ll} 0, & \text{Placebo} \\ 1, & \text{Drug} \end{array} \right\}$, $z_2 = \left\{ \begin{array}{ll} 0, & \text{Male} \\ 1, & \text{Female} \end{array} \right\}$ and $h(t|z_1, z_2) = h_0(t) \exp\{\beta_1 z_1 + \beta_2 z_2\}$. Let $h_0(t) = 1$, $\beta_1 = 1$ and $\beta_2 = 0.5$. Then

z_1	z_2	$h(t\|z)$
1	0	$e^1 \approx 2.72$
1	1	$e^{1.5} \approx 4.48$
0	0	$e^0 = 1$
0	1	$e^{0.5} \approx 1.64$

Here,

- $h_0(t) = 1$ represents the hazard for patient with covariates $(z_1, z_2) = (0, 0)$ (Male and Drug);
- $RR(\text{Drug : Placebo — Male}) = e^1/e^0 \approx 2.72$;
- $RR(\text{Drug : Placebo — Female}) = e^{1.5}/e^{0.5} \approx 2.72 = RR(\text{Drug : Placebo — Male})$;
- $RR(\text{Female : Male — Placebo}) = RR(\text{Female : Male — Drug}) \approx 1.64$.

In fact, when using this model, we have already assumed there is no interaction between gender and treatment. If we believe that there is interaction between gender and treatment, we have to consider the following Cox's model

$$h(t|z_1, z_2) = h_0(t) \exp\{\beta_1 z_1 + \beta_2 z_2 + \beta_3 (z_1 z_2)\}.$$

Then

- $RR(\text{Drug : Placebo — Male}) = \exp\{\beta_1\}/\exp\{0\}$;
- $RR(\text{Drug : Placebo — Female}) = \exp\{\beta_1 + \beta_2 + \beta_3\}/\exp\{\beta_2\}$;
- $RR(\text{Female : Male — Placebo}) = \exp\{\beta_2\}/\exp\{0\}$;
- $RR(\text{Female : Male — Drug}) = \exp\{\beta_1 + \beta_2 + \beta_3\}/\exp\{\beta_1\}$.

However, we still have

$$\frac{RR(\text{Drug : Placebo — Female})}{RR(\text{Drug : Placebo — Male})} = \frac{RR(\text{Female : Male — Drug})}{RR(\text{Female : Male — Placebo})}.$$

In fact, this equation should be hold no matter what model we assumed. This can be known through the definition of conditional probability.

Up to now, we discuss the partial likelihood function based on Cox's model by assuming there is no tie. If there are tie, recently we have four ways to adjust the partial likelihood function. Later, we use an example to verify the idea of these four adjustments.

If the time t is continuous, the probability to have tie data is zero. However, due to the measurement error, we may impossible to figure out the order of death time of the patients who have similar death time which in fact are different. For an example, suppose there are 3 patients died at the same time t. Let $R(t)$ be the risk set at time t, that is, the set of patients who are still alive just prior to time t. Denote the covariate for these three died patients be z_1, z_2 and z_3 respectively. Then there are 6 possible probabilities

$$
\begin{aligned}
p_1 &= \frac{\exp\{\beta z_1\}}{\sum_{j\in R(t)}\exp\{\beta z_j\}}\frac{\exp\{\beta z_2\}}{\sum_{j\in R(t)}\exp\{\beta z_j\}-\exp\{\beta z_1\}}\frac{\exp\{\beta z_3\}}{\sum_{j\in R(t)}\exp\{\beta z_j\}-\exp\{\beta z_1\}-\exp\{\beta z_2\}}\\
p_2 &= \frac{\exp\{\beta z_1\}}{\sum_{j\in R(t)}\exp\{\beta z_j\}}\frac{\exp\{\beta z_3\}}{\sum_{j\in R(t)}\exp\{\beta z_j\}-\exp\{\beta z_1\}}\frac{\exp\{\beta z_2\}}{\sum_{j\in R(t)}\exp\{\beta z_j\}-\exp\{\beta z_1\}-\exp\{\beta z_3\}}\\
p_3 &= \frac{\exp\{\beta z_2\}}{\sum_{j\in R(t)}\exp\{\beta z_j\}}\frac{\exp\{\beta z_1\}}{\sum_{j\in R(t)}\exp\{\beta z_j\}-\exp\{\beta z_2\}}\frac{\exp\{\beta z_3\}}{\sum_{j\in R(t)}\exp\{\beta z_j\}-\exp\{\beta z_1\}-\exp\{\beta z_2\}}\\
p_4 &= \frac{\exp\{\beta z_2\}}{\sum_{j\in R(t)}\exp\{\beta z_j\}}\frac{\exp\{\beta z_3\}}{\sum_{j\in R(t)}\exp\{\beta z_j\}-\exp\{\beta z_2\}}\frac{\exp\{\beta z_1\}}{\sum_{j\in R(t)}\exp\{\beta z_j\}-\exp\{\beta z_2\}-\exp\{\beta z_3\}}\\
p_5 &= \frac{\exp\{\beta z_3\}}{\sum_{j\in R(t)}\exp\{\beta z_j\}}\frac{\exp\{\beta z_1\}}{\sum_{j\in R(t)}\exp\{\beta z_j\}-\exp\{\beta z_3\}}\frac{\exp\{\beta z_2\}}{\sum_{j\in R(t)}\exp\{\beta z_j\}-\exp\{\beta z_3\}-\exp\{\beta z_1\}}\\
p_6 &= \frac{\exp\{\beta z_3\}}{\sum_{j\in R(t)}\exp\{\beta z_j\}}\frac{\exp\{\beta z_2\}}{\sum_{j\in R(t)}\exp\{\beta z_j\}-\exp\{\beta z_3\}}\frac{\exp\{\beta z_1\}}{\sum_{j\in R(t)}\exp\{\beta z_j\}-\exp\{\beta z_2\}-\exp\{\beta z_3\}}
\end{aligned}
$$

Let $s = z_1 + z_2 + z_3$, the four ways are

(1) Exact: $p = p_1 + \cdots + p_6$. (Note, if we assume this 6 possibilities have equal change to happen, the exact probability should be $p/6$. However, it does not matter if we use p or $p/6$ when considering partial likelihood technique since both of them will give an identical estimate of β. why?)

(2) Breslow's: Just use p_1 with replacement. Specifically

$$p = \frac{\exp\{\beta s\}}{[\sum_{j\in R(t)}\exp\{\beta z_j\}]^3}.$$

(3) Effron's: Between exact and Breslow, use p_1 without replacement. Specifically

$$p = \frac{\exp\{\beta s\}}{[\sum_{j\in R(t)}\exp\{\beta z_j\}][\sum_{j\in R(t)}\exp\{\beta z_j\}-\frac{1}{3}\exp\{\beta s\}][\sum_{j\in R(t)}\exp\{\beta z_j\}-\frac{2}{3}\exp\{\beta s\}]}$$

(4) Discrete:

$$p = \frac{\exp\{\beta s\}}{\sum_{j,j',j''\in R(t), j\neq j'\neq j''}\exp\{\beta(z_j + z_{j'} + z_{j''}\}}.$$

In addition, SAS has provided these four ways to estimate β in PROC PHREG.

Stratification:

In many application, we want to estimate the RR for treatment adjusted for other covariates such as age. Suppose there are three age groups, child, young and old group. Let $z_1 = \left\{ \begin{array}{ll} 1, & \text{if young} \\ 0, & \text{o.w} \end{array} \right\}$ and $z_2 = \left\{ \begin{array}{ll} 1, & \text{if old} \\ 0, & \text{o.w} \end{array} \right\}$. It may possible that

$$h(t|treat, z_1, z_2) = \left\{ \begin{array}{ll} h_1(t)\exp\{\beta treat\}, & \text{if child} \\ h_2(t)\exp\{\beta treat\}, & \text{if yound} \\ h_3(t)\exp\{\beta treat\}, & \text{if old} \end{array} \right\},$$

however, $h_1(t), h_2(t)$ and $h_3(t)$ are not proportional to each other. That is, given the age group, the hazard function of treat and non-treat groups are proportional to each other, however, the proportional assumption fail within age groups. Therefore, we can not assume

$$h(t|treat, z_1, z_2) = h_0(t)\exp\{\beta treat + \beta_1 z_1 + \beta_2 z_2\}.$$

In this case, we can consider the stratify Cox's proportional hazards model.

(1) Assume proportional hazards fail within age and the hazard ratio in different age groups are different. Then

$$h(t|treat, z_1, z_2) = \left\{ \begin{array}{ll} h_1(t)\exp\{\beta_1 treat\}, & \text{if child} \\ h_2(t)\exp\{\beta_2 treat\}, & \text{if yound} \\ h_3(t)\exp\{\beta_3 treat\}, & \text{if old} \end{array} \right\}.$$

It is identical to running three Cox's models with three sub-observations respectively.

(2) Assume proportional hazards fail within age and the hazard ratio in different age groups are the same. Then

$$h(t|treat, z_1, z_2) = \left\{ \begin{array}{ll} h_1(t)\exp\{\beta treat\}, & \text{if child} \\ h_2(t)\exp\{\beta treat\}, & \text{if yound} \\ h_3(t)\exp\{\beta treat\}, & \text{if old} \end{array} \right\}.$$

(3) Assume no age effect (i.e. $h_1 = h_2 = h_3$) but the hazard ratio in different age groups are different. Then

$$h(t|treat, z_1, z_2) = h_0(t)\exp\{\beta_1 treat + \beta_2 treat * z_1 + \beta_3 treat * z_2\}.$$

(4) Assume no age effect and the hazard ratio in different age groups are the same. Then

$$h(t|treat, z_1, z_2) = h_0(t) \exp\{\beta treat\}.$$

It is exactly the regular Cox's model withoud adjusted by age.

How to estimate $H(t) = \int_0^t h_0(s)ds$ based onf $h(t|z) = h_0(t) \exp\{\beta z\}$?