PH Model Estimation

Assume

- Observation: $(y_1, \delta_1, x_1), \cdots, (y_n, \delta_n, x_n)$
- Independent censoring: $T_i \perp \delta_i$ (the actual failure time is independent of the censoring indicator)
- Uncensored failure times are distinct (no tie), and ordered as $y_{(1)} < \cdots < y_{(k)}$ (note $k \leq n$)

Then, we can use partial likelihood approach to estimate β in the PH model

Partial Likelihood

Full likelihood of $(y_1, \delta_1), \cdots, (y_n, \delta_n)$:

$$L = \prod_{i=1}^{n} f(y_i \mid x_i)^{\delta_i} S(y_i \mid x_i)^{1-\delta_i}$$

which involves the nonparametric baseline hazard function $h_0(t)$.

We are more interested in β than $h_0(t)$. Thus, we use the partial likelihood (the probability of having $x_{(i)}$ failed at $y_{(i)}$, given multiple observations are at risk at $y_{(i)}$ and exactly one fails at $y_{(i)}$). The partial likelihood is

$$L = \prod_{i'=1}^{k} \frac{\exp(x_{(i')}\beta)}{\sum_{j \in R_{(i')}} \exp(x_j \beta)}$$

where $R_{(i')}$ is the collection of sample indices (censored and uncensored) at risk at $y_{(i')}$.

Estimation of Baseline Hazard Function

Now we know how to estimate β, but what about $h_0(t)$?

- Breslow estimator
- Piece-wise constant baseline hazard function
- Requires large sample size
- Much less popular than β

Time-Dependent Covariate

If the value of the covariate change over time, the variable is called time dependent covariate.

- Age at survival time T
- treatment plan changes over time
- drug dosage level changes over time
- biomarker value changes over time

Time-dependent covariates are represented as

$$x_i(t) = (x_{i1}(t), \cdots, x_{ip}(t))$$

The PH model becomes

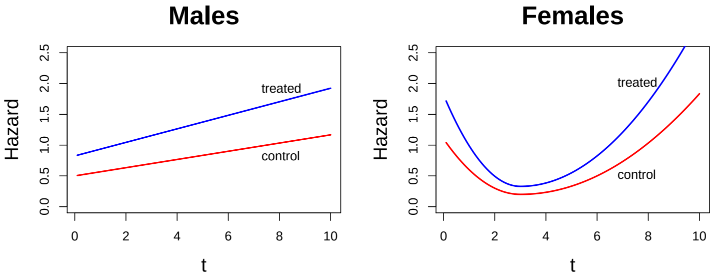
$$h_i(t) = h_0(t) \exp(x_i(t)\beta)$$

- β can still be estimated using the partial likelihood approach.
- β needs to be interpreted at a given time point.

Stratified PH Model

The proportional hazards assumption might not hold for all covariates, but it may be true within each level of a (stratification) covariate.

For example, HR for treated vs. control is constant for females and males respectively, but HR for female vs. male may not be constant. Also, the shapes of baseline hazard functions across strata can be different.



- We need a stratified PH model:

$$h_{im}(t) = h_{0m}(t)e^{x_i\beta}$$

$$h_{if}(t) = h_{0f}(t)e^{x_i\beta}$$

- The model requires proportional hazards assumption for individuals within the same stratum.
- Baseline hazard function can be different for males and females.
- For other covariates in x_i (e.g., treatment), the hazard ratio is constant across strata.
- Use partial likelihood to estimate β.

Example

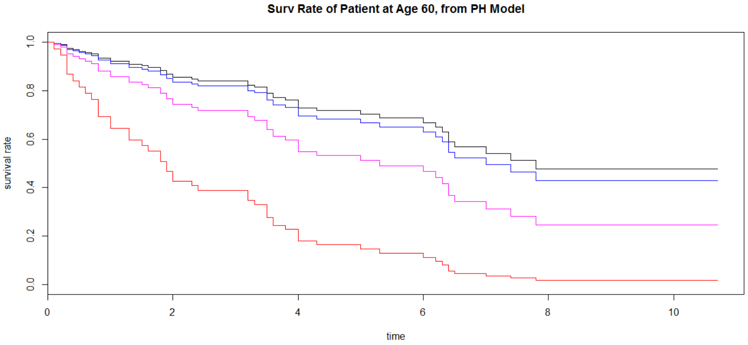
Larynx cancer:

- 90 males with larynx cancer.
- Interested in time to death.
- Covariates include different stages of disease (multiple covariates or strata) and age at diagnosis (time-independent).
- Question: how does the baseline age and stage affect the survival outcome?

Model:

$$h(t) = h_0(t)\exp(x_1\beta_1 + x_2\beta_2 + x_3\beta_3 + age * \beta_4)$$

where x_1, x_2, x_3 are indicators of stage II, III, and IV, respectively.



Model:

$$h_I(t) = h_{0I}(t)\exp(age * \beta), \quad h_{II}(t) = h_{0II}(t)\exp(age * \beta)$$
$$h_{III}(t) = h_{0III}(t)\exp(age * \beta), \quad h_{IV}(t) = h_{0IV}(t)\exp(age * \beta)$$