Biostatistics (MATH11230)

Vanda Inácio

University of Edinburgh

Semester 1, 2021/2022

Proportional hazards model

Adjusting survival curves

↪ Since the baseline hazard is considered a nuisance parameter and it is not estimated, Cox semiparametric proportional hazards model cannot be used directly to obtain an estimator of the survival curves of subjects with specific covariate values.

↪ The baseline hazard has to be estimated and one proposal is to extend the Nelson–Aalen estimator for the case of risk factors:

$$\widehat{H}_0(t) = \sum_{j:t_j \leq t} \frac{d_j}{\sum_{l \in R(t_j)} \exp(\mathbf{x}_l' \widehat{\boldsymbol{\beta}})},$$

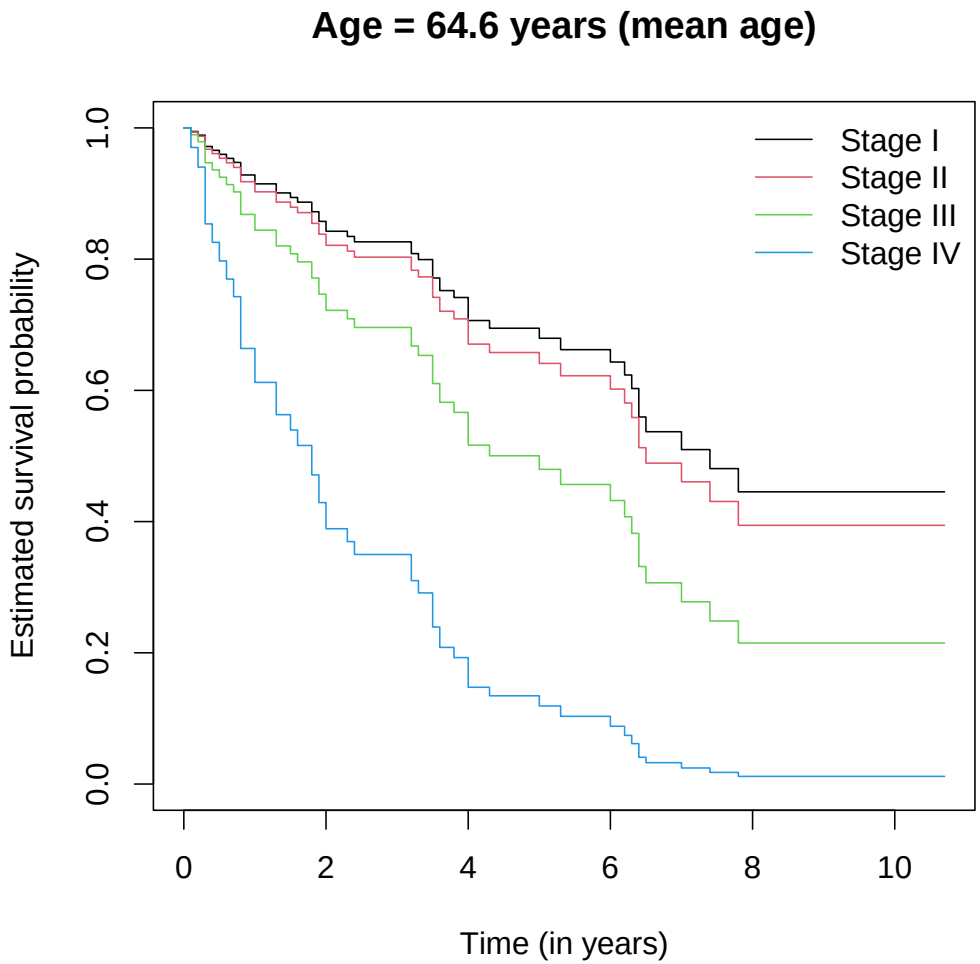
with $t_1, \ldots, t_J$ being the ordered distinct event times and d_j the number of events at time t_j.

↪ The survival curve for subjects with covariate values $\mathbf{x}_i$ is then estimated by

$$S(t \mid \mathbf{x}_i) = \left[\widehat{S}_0(t)\right]^{\exp(\mathbf{x}_i' \widehat{\boldsymbol{\beta}})}, \quad \widehat{S}_0(t) = \exp\{-\widehat{H}_0(t)\}.$$

Proportional hazards model

Adjusting survival curves



Proportional hazards model

Checking the proportional hazards assumption

- Over the last decades, Cox proportional hazards model became, possibly, the most popular regression model for time to event/survival data and is widely used in medicine and beyond.
- An advantage of the proportional hazards model is that it summarises the effect of each risk factor by a single summary measure, the hazard ratio.
- For example, in a clinical trial with a time to event endpoint, it is common to express the treatment effect as an hazard ratio.
- However, the proportional hazards assumption is not always realistic and in particular, in medicine, it has been challenged over the last years by the emergence of new types of treatments having different mechanisms of action.
- In oncology, for example, it is expected that immunotherapy has a delayed treatment effect and comparison with a more standard treatment in a clinical trial will most probably lead to a violation of the proportional hazards assumption.

Proportional hazards model

Checking the proportional hazards assumption: the log cumulative hazard plot

↪ The proportional hazards model is given by

$$h(t \mid \mathbf{x}_i) = h_0(t) \exp(\mathbf{x}_i' \boldsymbol{\beta}).$$

↪ Integrating both sides from 0 up to t, yields

$$\int_0^t h(u \mid \mathbf{x}_i) \mathrm{d}u = \int_0^t h_0(u) \exp(\mathbf{x}_i' \boldsymbol{\beta}) \mathrm{d}u$$
$$\Rightarrow H(t \mid \mathbf{x}_i) = H_0(t) \exp(\mathbf{x}_i' \boldsymbol{\beta}),$$

where $H(t \mid \mathbf{x}_i)$ and $H_0(t)$ are the cumulative hazard functions.

Proportional hazards model

Checking the proportional hazards assumption: the log cumulative hazard plot

↪ Taking logarithms of each side of this equation, we get

$$\log H(t \mid \mathbf{x}_i) = \log H_0(t) + \mathbf{x}_i'\beta,$$

or, equivalently,

$$\log\{-\log S(t \mid \mathbf{x}_i)\} = \log\{-\log S_0(t)\} + \mathbf{x}_i'\beta.$$

↪ It thus follows that differences in the log cumulative hazard functions do not depend on time.

↪ This means that if the log cumulative hazard functions for individuals with different values of their explanatory variables are plotted against time, the curves so formed will be parallel if the proportional hazards model is valid.

↪ This approach only requires estimating $H(t \mid \mathbf{x}_i)$ or $S(t \mid \mathbf{x}_i)$ nonparametrically, for example, via the Nelson–Aalen or the Kaplan–Meier estimator, for the groups of individuals defined by the different values of the risk factor(s) under study.

Proportional hazards model

Checking the proportional hazards assumption: the log cumulative hazard plot

- ↪ This approach is most useful when we have a limited number of (categorical) risk factors and a sufficient number of observations per each level of the risk factors.
- ↪ The resulting estimated curves will usually be step functions, due to the nature of the underlying estimators of the survival/cumulative hazard functions, and might be rather poorly estimated for the levels of the risk factor with less observations.
- ↪ To decide whether these curves are indeed parallel will obviously always be subject to some subjectivity.
- ↪ A commonly adopted attitude is to assume that the proportional hazards assumption is satisfied unless these curves show a strong deviation from parallel curves.
- ↪ Continuous risk factors can also be considered but must first be categorised. It is usually recommended to consider a reasonably low number of categories, keeping sufficient observations in each category.

Proportional hazards model

Checking the proportional hazards assumption: the log cumulative hazard plot

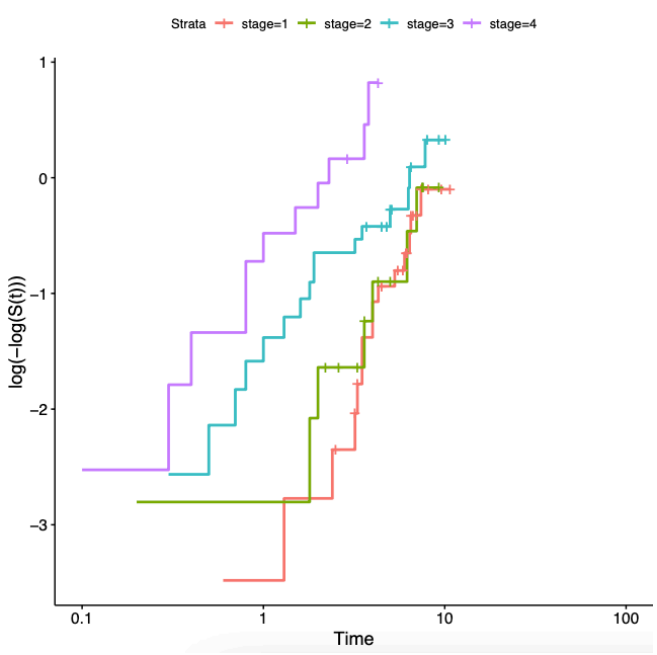
- ↪ While in theory this approach can be used to evaluate the proportional hazards assumption for several risk factors at the same time, inspecting all possible combinations of the various levels of each the risk factors easily gets impractical.

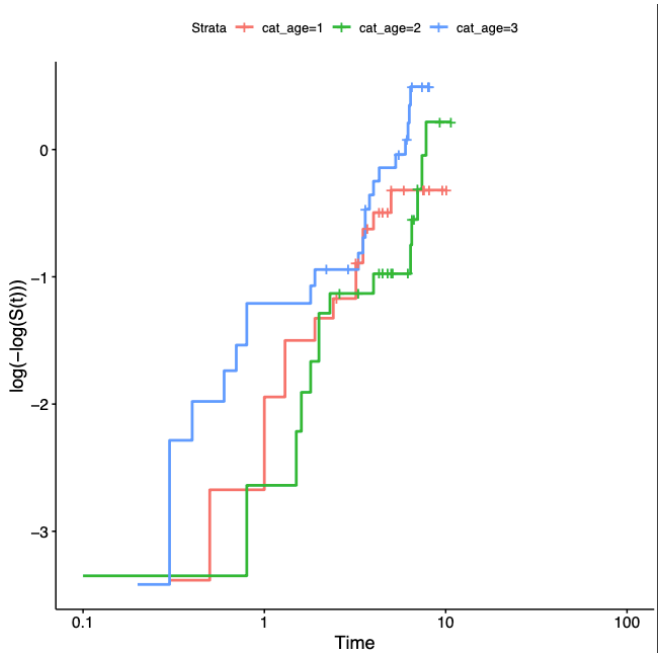
- ↪ Further, when stratifying by several risk factors, the number of observations in each combination may be so low that it prevents accurate estimation of the corresponding survival/cumulative hazard function.

Proportional hazards model

Checking the proportional hazards assumption: the log cumulative hazard plot

↪ Investigating if the proportional hazards assumption is valid for the variables stage and age in the larynx cancer example (age has been categorised according to the 0.33 and 0.66 quantiles).





Proportional hazards model

Checking the proportional hazards assumption: Schoenfeld residuals

- Another graphical option is to use the Schoenfeld residuals, also known as partial residuals.
- For the kth covariate, $k = 1, \ldots, p$, and the ith individual, the Schoenfeld residual represents the difference between the observed value of x_{ik} and its conditional expectation given the risk set $R(y_i)$

$$r_{ik} = \delta_i(x_{ik} - \hat{a}_{ik}), \quad \hat{a}_{ik} = \frac{\sum_{l \in R(y_i)} x_{lk} \exp(\mathbf{x}_l' \widehat{\boldsymbol{\beta}})}{\sum_{l \in R(y_i)} \exp(\mathbf{x}_l' \widehat{\boldsymbol{\beta}})}.$$

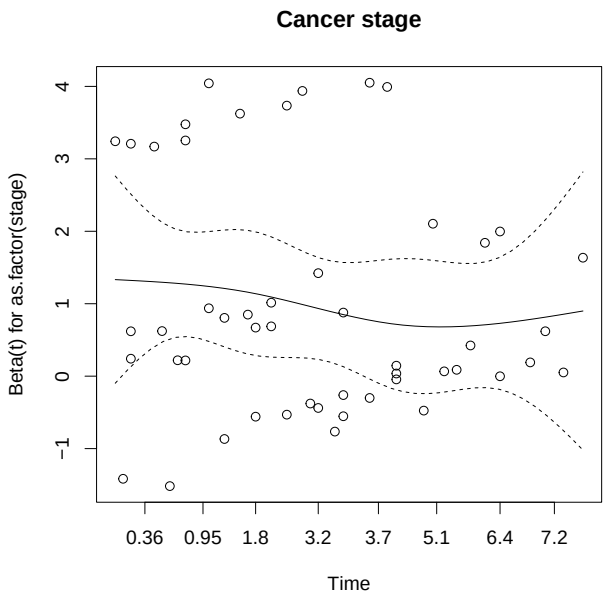
- Non-zero values of these residuals only arise for uncensored observations.
- Moreover, if the largest observation in a sample of survival times is uncensored, the value of $\hat{a}_{ik}$ for that observation will be equal to x_{ik} and so $r_{ik} = 0$.

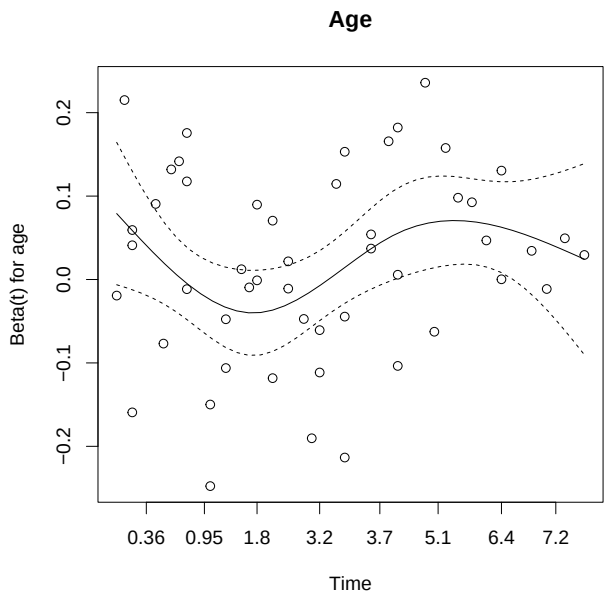
Proportional hazards model

Checking the proportional hazards assumption: Schoenfeld residuals

- ↪ It has been shown that if the proportional hazards assumption is valid, Schoenfeld residuals should be flat and centred about zero.

- ↪ It has also been demonstrated that these residuals have to be independent of the time and so if we represent them ranked by its event time, this plot must not show any pattern.





Proportional hazards model

Goodness-of-fit test

- A second approach for assessing the PH assumption involves goodness-of-fit tests.
- To this end, different test have been proposed in the literature (Grambsch and Therneau 1994).
- We focus on the test proposed by Harrell (1986), a variation of a test originally proposed by Schoenfeld (1982). This is a test of correlation between the Schoenfeld residuals and event time time.
- A correlation of zero indicates that the model meet the proportional hazards assumption (the null hypothesis).

Proportional hazards model

Non-Proportional Hazards... and now what?

- A 'modest' violation of the proportional hazards assumption may not make a big difference for large datasets.
- What if the nonproportionality is large and real?
- Risk factors with nonproportional effects may be incorporated into the model as stratification factors rather than explanatory variables.
- The idea is to consider a categorical variable defining the 'strata' in the population and to consider a different (unspecified) baseline hazard function in each strata.
- So, the hazard for individual i in strata j, $i = 1, \ldots, n_j$ and $j = 1, \ldots, J$, with risk factor values x_{ij}, is now given by
$$h_{ij}(t) = h_{0j}(t) \exp(\mathbf{x}'_{ij}\beta).$$
- The strata can, for example, be defined based on the gender.
- While this may be convenient for some applications to overcome the nonproportional hazards issue, we loose all the information about the stratification factor in the sense that this model does not provide information on the effect of the stratification variable.